1

Haskins Laboratories Status Report on

Speech Research

SR-109/110
JANUARY-JUNE 1992

Haskins
Laboratories
Status Report on

Speech Research

SR-109/110
JANUARY-JUNE 1992

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Michael Studdert-Kennedy

Production
Yvonne Manning
Fawn Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8820099

National Institute on Deafness and Other Communication Disorders

Grant DC 00121	Grant DC 00865
Grant DC 00183	Grant DC 01147
Grant DC 00403	Grant DC 00044
Grant DC 00016	Grant DC 00825
Grant DC 00594	Grant DC 01247

Investigators

Arthur Abramson*
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Lois G. Dreyer*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Carol A. Fowler*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Berit Halvorsen†
Katherine S. Harris*
John Hogden
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Ilya Leykin*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Kiyoshi Oshima‡
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Hyla Rubin*
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Rudolph Sock§
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Lisa Fresa
Vincent Gulisano
Donald Hailey
Maura Herlihy
Raymond C. Huey*
Marion MacEachron
Yvonne Manning
Joan Martinez
William P. Scully
Fawn Zefang Wang
Edward R. Wiley

Students*

Melanie Campbell
Sandra Chiang
Margaret Hall Dunn
Terri Erwin
Joseph Kalinowski
Laura Koenig
Betty Kollia
Simon Levy
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Christine Romano
Joaquin Romero
Maria Rosetti
Dorothy Ross
Arlyne Russo
Michelle Sancier
Jeffrey Shaw
Sonya Sheffert
Caroline Smith
Brenda Stone
Mark Tiede
Qi Wang
Yi Xu
Elizabeth Zsiga

*Part-time
†Visiting from University of Bergen, Norway
‡Visiting from University of Tokyo, Japan
§Visiting from University of Grenoble, France

Contents

Phonological and Articulatory Characteristics of Spoken Language
Carol A. Fowler 1

Characteristics of Speech as a Motor Control System
Vincent L. Gracco 13

Sensorimotor Mechanisms in Speech Motor Control
Vincent L. Gracco 27

Analysis of Speech Movements: Practical Considerations and Clinical Application
Vincent L. Gracco 45

Reiterant Speech as a Test of Nonnative Speakers' Mastery of the Timing of French
Andrea G. Levitt 59

Syllable-internal Structure and the Sonority Hierarchy: Differential Evidence from Lexical Decision, Naming, and Reading
Andrea Levitt, Alice F. Healy, and David W. Fendrich 73

Effects of Phonological and Phonetic Factors on Cross-Language Perception
Catherine T. Best and Winifred Strange 89

Plausibility, Parsimony, and Theories of Speech
Alvin M. Liberman 109

The Relation of Speech to Reading and Writing
Alvin M. Liberman 119

Linguistic Awareness and Orthographic Form
Ignatius G. Mattingly 129

The Effects of Aging and First Grade School on the Development of Phonological Awareness
Shlomo Bentin, Ronen Hammer, and Sorel Cahan 141

Bi-alphabetism and the Design of a Reading Mechanism
Laurie Beth Feldman 147

Morphological Analysis of Disrupted Morphemes: Evidence from Hebrew
Laurie Beth Feldman and Shlomo Bentin 163

Phonetic Recoding of Print and Its Effect on the Detection of Concurrent Speech in Amplitude Modulated Noise
Ram Frost 179

Processing Phonological and Semantic Ambiguity:
Evidence From Semantic Priming at Different SOAs
Ram Frost and Shlomo Bentin 191

Attention Mechanisms Mediate the Syntactic Priming Effect in Auditory
Word Identification
Avital Deutsch and Shlomo Bentin 205

Starting on the Right Foot. A review of Marilyn Jager Adams'
Beginning to Read: Thinking and Learning about Print
Donald Shankweiler 221

Null Subject vs. Null Object: Some Evidence from the Acquisition of Chinese
and English
Qi Wang, Diane Lillo-Martin, Catherine T. Best, and Andrea Levitt 225

Amplitude as a Cue to Word-initial Consonant Length: Pattani Malay
Arthur S. Abramson 251

Tone Splits and Voicing Shifts in Thai: Phonetic Plausibility
Arthur S. Abramson and Donna M. Erickson 255

A Constraint on the Expressive Timing of a Melodic Gesture:
Evidence from Performance and Aesthetic Judgment
Bruno H. Repp 263

A Review of Carol L. Krumhansl's *Cognitive Foundations of Musical Pitch*
Bruno H. Repp 277

Toward an Emancipation of the "Weaker Sense"
Bruno H. Repp 283

Appendix 289

Haskins Laboratories Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 1-12

Phonological and Articulatory Characteristics of Spoken Language*

Carol A. Fowler†

1. INTRODUCTION

Speaking may be our most impressive motor skill. We speak rapidly, and production of each word involves intricate sequencing and temporal interleaving of gestures for the component, ordered consonants and vowels of the word. The problem of understanding speech production at this level is that of understanding how speakers accomplish the feat of fluent consonant and vowel production. Solving that problem involves solving another one, however. It is to understand what speaking is essentially. That is, it is to understand how a series of complicated actions of a vocal tract can serve to convey a message composed of rulefully-patterned symbols to members of a language community. In fact, the kind of solution an investigator seeks to the problem of understanding how vocal-tract actions are executed depends on how the investigator looks at the relation between vocal-tract action and the linguistic message itself.

Phonology is traditionally seen as the discipline that concerns itself with the building blocks of linguistic messages. It is the study of the structure of sound inventories of languages and of the participation of sounds in rules or processes. Phonetics, in contrast, concerns speech sounds as produced and perceived. Two extreme positions on the relationship between phonological messages and phonetic realizations are represented in the literature. One holds that the primary home for linguistic symbols, including phonological ones, is the human mind, itself housed in the human brain. The second holds that their primary home is the human vocal tract.

Consider the first position and the conceptualization of speech production to which it leads. For at least two reasons, the vocal tract is rejected as a natural home for phonological segments of the language. A philosophical reason is that phonemes are not the kinds of things that can occur or exist outside the mind. They are ideas or concepts without real-world actualization. Articulatory gestures or their acoustic consequences can serve as cues to phonological segments, but they cannot be phonological segments.

> "[Segments] are *abstractions*. They are the end result of complex perceptual and cognitive processes in the listener's brain" (Repp 1981, 1462)

> "Phonological representation is concerned with speakers' implicit knowledge, that is with information in the mind...[Phonetic] representation...is not cognitive because it concerns events in the world rather than events in the mind." (Pierrehumbert, in press)

A practical reason why phonological segments cannot occur in the vocal tract is that linguistic symbols have other properties, aside from being covert kinds of things, that preclude the vocal tract from representing them veridically or even analogically. In particular, a central and important fact about language is that its messages are composed of discrete symbols. Phonological segments are discrete in the sense that they do not overlap and blend. Moreover, until recently, they have been represented in linguistic theories as if they were composed of lists of coextensive (and by implication, cotemporal) features (cf. Chomsky & Halle, 1968). The features themselves described static postures of the vocal tract or their acoustic consequences; accordingly, the feature lists of a word described a succession of vocal-tract or acoustic snapshots. The vocal-tract actions that somehow convey a message to a listener have none of those properties. Actions associable with a given consonant or vowel do overlap and do appear to blend with actions of neighbors. Actions identifiable with the component features of a consonant or vowel are not cotemporal. Finally, fun-

damental units of articulation appear to be actions, not postures; accordingly, time is intrinsic to speech, rather than extrinsic as it is to the linguistic message. One interpretation of these mismatches is that they reflect the mismatch between the ideal of linguistic competence and the degraded physical reality of linguistic vocal performance; the latter necessarily is a considerable distortion of the former due to the limitations of mechanico-inertial systems. This way of looking at speech production promotes development of a kind of theory of the "how" of speech production that have been termed translation theories (Fowler, Rubin, & Remez et al., 1980). The mismatch between the character of the planned message, presumably a sequence of linguistic symbols, and of its physical, phonetic, realization requires a translation over stages of processing out of the ideal, mental, domain of the plan into the real, physical-nonmental, domain of a vocal tract.

The other extreme perspective on the nature of speaking is that consonants and vowels are actions of the vocal tract that have linguistic, including phonological, significance in a language community. They are, certainly, psychological actions that require knowledge about them to be performed. However, the knowledge is not a superior "ideal" that the actions cannot implement; rather, the knowledge is about the actions, derived from perceptual and articulatory experience with them. From this perspective, the mismatch between linguistic segments and articulation described above is apparent rather than real. It is the product of three kinds of error: 1. a mistaken ascription of primacy to linguistic knowledge (competence) over linguistic activity (performance); 2. an incorrect characterization of phonological segments in linguistic theory; 3. an incorrect characterization of the vocal tract actions of speech production. As to the first "error," the argument is that we treat language differently from other human creations when we decide that its components exist only in the mind. Other human creations include, for example, automobiles, baseball games and musical pieces. Automobiles definitely exist in the world and so do baseball games and musical pieces when they are played. What is in the mind of those who know about automobiles, baseball and a musical piece, is only what they know about those things; it is not the things themselves. If linguistic concepts are like these other concepts, they are knowledge about real-world objects or events; the events have a psychological nature--in this case, they are actions of the vocal tract, identified as phonological segments. If the phonology in the mind of a language user is what a the user knows about the actions that implement a linguistic message, then there need be no mismatch between knowledge and action. If a phonological theory ascribes properties to phonological segments as known that are impossible to realize in vocal-tract action, then the first hypothesis should be that the theory is wrong, not that vocal-tract action distorts components of linguistic competence. If descriptions of vocal-tract actions include properties, such as coarticulatory blending, that would distort the phonological message, then the first hypothesis should be that the descriptions are wrong. From this perspective, an important aim is to work on development of a phonology that does not ascribe properties to phonological segments that are unproduceable as vocal-tract action (cf. Browman & Goldstein, 1986; Browman & Goldstein, 1989). A second aim is to find a perspective on vocal-tract action from which macroscopic order is evident that conforms to the phonological structure of spoken utterances (e.g., Fowler, Rubin, & Remez, et al., 1980; Fowler, in press; Saltzman, 1986; Saltzman & Munhall, 1989).

This theoretical perspective promotes a theory of speech production different from a translation theory as outlined earlier. Speech production does not involve a translation out of an ideal, mental domain into a physical, nonmental, domain. Rather, the plan for a sequence of phonological segments, physically instantiated in the brain, replicates itself in a new physical medium, the moving vocal tract. A speech plan, in some way, brings about vocal-tract actions having linguistic significance.

In the remainder of this chapter, I pursue the different outlooks on a central aspect of speech production, coarticulation, that these different theoretical perspectives promote. I then consider the implications of our understanding of coarticulation for understanding another central aspect of speech production: the coordinated actions of the vocal tract that constitute token phonological segments.

2. TWO PERSPECTIVES ON COARTICULATION

All sources of evidence regarding speech production, whether they are acoustic or articulatory, provide the same general picture of context-sensitivity in speech production. An acoustic signal displayed spectrographically or as a waveform, for example, can be divided into phonological-segment sized regions (e.g., Klatt

1973) by identifying acoustic properties that are more strongly associated with one particular segment of an utterance than with others. For example a stop burst can be assigned to a stop in a stop-vowel utterance, and the following voiced formants can be assigned to the vowel. Even so, however, the display has not thereby been partitioned into phonological segments or even into their acoustic consequences. This is so, in part, because there may be no obvious place to locate a boundary separating the acoustic consequences of one phoneme from those of another. For example, the voiceless formant transitions following a voiceless stop consonant in a consonant-vowel sequence belong with the stop, because they are voiceless, but they also belong with the vowel, because formants are characteristic of vowels and other sonorants (see, e.g., Peterson & Lehiste, 1960). Indeed, generally, there are no boundaries *between* segments so that a partition leaves all and only the acoustic information for one segment on one side of the boundary and all and only that for another segment on the other side (cf. Fant & Lindblom, 1961). Moreover, the overlap is not only in a potential boundary region. Spectral analysis of the signal well within a domain associated with a particular phonetic segment—well within the frication region for a fricative or within the steady-state formants, if any, of a vowel, for example—is likely to reveal influences of context. (I will use the term "domain" to refer to the temporal region in which the features of a segment dominate in articulation or in the acoustic signal. The domain does not include the whole articulatory extent of a segment or the whole region in which it influences the acoustic signal, but only the region in which it is dominant; see also Löfqvist, 1990.)

Examination of the articulatory behaviors that give rise to acoustic speech signals reveals a compatible picture. Articulatory movements can be found that are identifiable with one of the phonetic segments in an utterance—movement toward bilabial closure in a bV sequence, for example. In addition, boundaries can be located around that movement. In the example, a boundary may be located where closing is first detectable and another at the point of release of the closure. Once again, however, the boundaries are not boundaries between phonological segments or their articulatory consequences so that all and only movements associated with /b/ occur within the boundaries and movements associated with other segments fall outside the boundaries. During the closing and closure gestures, the tongue body will be conforming itself to the requirements of the following vowel (e.g., Öhman 1966), and once again, the movements within the boundaries are context-sensitive. For example, the jaw moves to a higher point of maximum closing for /b/ followed by a high than a low vowel (Keating, Lindblom, and Lubker, cited in Keating, 1985).

Sources of context sensitivity are bidirectional. Effects of earlier segments in the string extending beyond their domains of prominence are termed "left-to-right," "perseverative" or "carryover" effects. Effects of later segments are called "right-to-left" or "anticipatory." Estimates of the coarticulatory field—that is, the interval of time or the number of segmental domains affected by a segment in either direction—vary considerably, but may be quite large. For example, Magen (1989) reports anticipatory effects of V_3 on $V_1C\partial CV_3$ sequences in English. While some part of the carryover influences can been ascribed to inertial properties of the vocal tract and to its inability instantaneously to adopt a characteristic posture for one phonological segment without exhibiting transitional movements between the postures, anticipatory coarticulation cannot have that explanation, and carryover effects are sometimes more extensive than can be realistically ascribed to these mechanical factors (Daniloff & Hammarberg, 1973). These considerations have suggested to many investigators that coarticulation is planned. Generally accounts of coarticulation diverge along the theoretical lines distinguished in the introduction.

2.1 Coarticulation as assimilation by feature spreading

In a translation theory, coarticulation serves an important function of, indeed, translating a planned symbol string into a form more compatible with the capabilities of vocal-tract action. (The role of phonetic rules generally, according to Keating (1988a), is to make the linguistic representation "more physical.")

One example of a theory in which coarticulation serves that function is that of Daniloff and Hammarberg (1973). Daniloff and Hammarberg described the phonological segments that serve as "input" in a plan to speak as "canonical forms"—that is, "invariant, ideal, uncoarticulated forms"—the phonological types of a linguistic theory. These forms undergo "articulatory encoding" to tailor them to the vocal tract. The encoding processes include application of context-sensitive rules of

feature spreading. An example they provide of such a rule is one that spreads a rounding feature from a vowel to a preceding /ʃ/: ʃ --> ʃʷ / __ [+round, +V]. By this rule, the /ʃ/ in "shoe," for example, acquires the feature [+round] from its context, a following rounded vowel. Generally (following Henke 1966), rules cause a feature to spread in an anticipatory direction to any phonetic segment that is "unspecified" for that feature. Feature values in phonological theory generally are binary, and a segment may be "specified" for a feature by having either a "+" or a "-" value of that feature. (Accordingly, a rounded vowel is [+round] while an unrounded vowel is [-round].) To count as an instance of a segment specified for some feature value, a token occurrence of the segment must have the appropriate feature value; changing the value may change one segment into another and hence, in a sequence of phonemes, may change one word into another. These feature values thereby serve a "contrastive" function in the language. At least hypothetically, the contrastive feature values cannot be changed by a feature-spreading coarticulatory rule. However, some features are irrelevant to the identification of some segments. For example, in English, rounding is not contrastive for consonants; accordingly, making a consonant rounded does not change it from one consonant of English to another. Consonants are said to be "unspecified" for rounding, and they are subject to coarticulatory rules of feature spreading.

Evidence compatible with the feature- spreading theory includes findings (or, perhaps, interpretations of findings; see 2.2) that lip rounding anticipates a rounded vowel across any number of preceding consonants (e.g., Daniloff & Moll, 1968); (Benguerel & Cowan, 1974) and that nasality anticipates a nasal consonant across any number of vowels uninterrupted by oral consonants (Moll & Daniloff, 1971).

The simple characterization of coarticulation fails in several ways. One is that the coarticulatory field very often does not respect boundaries drawn between segments. That is, the hypothesis of feature spreading as the sole source of coarticulation predicts that the spread feature should be uniformly present throughout the production of the segment—at least to the same extent that other features of the segment are present, but that is generally not the case (e.g., Benguerel & Cowan, 1974; Krakow, 1989). Second, the magnitude of effects of ostensibly spread features is gradient rather than categorical. For example, Manuel and Krakow (1984) found that a following (front, high) vowel /e/, raises and fronts following (low, back) vowel /a/, but (front, high) /i/ raises it even more. Likewise, Marchal (1988) reported graded effects of one stop consonant on another in /kt/ sequences that suggested varying degrees of coarticulatory overlap between them. A third problem is that coarticulatory influences may affect realizations of specified features. In Marchal's findings, just cited, coarticulatory influences occur between stops specified for different places of articulation. A final problem relates to the idea of underspecification. The problem here is that segments considered to be unspecified for a feature involving some articulator—say, rounding and the lips (in English, consonants) or nasality and the velum (in English, vowels)—are not wholly neutral with respect to the demands they make on the articulator. Some consonants are associated with rounding movements of the lips (for example, /l/, /r/ and /s/ and /ʃ/ (Bell-Berti & Harris, 1982; Delattre & Freeman, 1968; Leidner, 1973). Compatibly, vowels, ostensibly unspecified for nasality are associated with characteristic postures of the velum (Bell-Berti, 1980; Moll, 1962). Despite their not being wholly unspecified in terms of articulatory control, they are subject to coarticulatory influences from specified neighbors and they coarticulate with neighbors. For example, the different velum heights associated with vowels of different heights both influence velum height for neighboring consonants and they are recipients of coarticulatory influences from nasal consonants (Bell-Berti, 1980). Accordingly, in contrast to the feature-spreading account of coarticulation, coarticulatory influences occur in the absence of any linguistic features to spread.

Recently, Keating (1988 a,b) has proposed an alternative account of specification and its role in coarticulation that preserves the idea of coarticulation as a participant in a translation from the mental to the physical domain of talking. She proposes that coarticulation includes processes at two levels at least, one phonological and one phonetic. At the phonological level, coarticulation is assimilatory feature spreading. Since Keating's focus has been on phonetic coarticulation, she simply alludes to this type of coarticulation without providing an example. However, a possible example is provided by Daniloff and Hammarberg (1973). They point out that in the word "width," there is, apparently, a spreading of the interdental place of articulation of /q/ to /d/ (which, by the way, is *specified* for a different place of articulation; however, in this case the feature change does not yield a different phoneme of English). As for phonetic

coarticulation, Keating proposes a "targets and connections" model. In the model, phonetic segments are associated with characteristic targets, and segments are sequenced by interpolating between successive targets. A novel aspect of Keating's idea of targets, however (but see Manuel 1987, for a similar idea), is that the targets are regions ("windows"), rather than fixed postures. Windows differ in their widths, and a target's instantiation within its window will depend on its neighbors in that the speaker will generally select the most efficient path from segment to segment that passes through each target region. The idea of target windows replaces the idea of underspecification as "categorial" with a gradient version. A segment with the narrowest possible window for some feature is "specified" for that feature value; one with the widest possible window for a feature is unspecified. However, most segments have intermediate target window sizes for their component features. Vowels have wider windows for velum height than do nasal and oral consonants, but the window is not as wide as possible. Accordingly, a vowel's window region does affect the articulatory path through the target window of neighboring segments.

This model handles the data of coarticulation considerably better than does the feature spreading model of Daniloff and Hammarberg (1973); yet it preserves the idea of coarticulation as among the processes that make the planned utterance "more physical." The targets and connections model is not obviously consistent with all of the data, however. In particular, one finding that the model does not seem to handle well is the ubiquity of coarticulatory fields that extend beyond immediate neighbors. The targets and connections idea explains how contiguous segments can be produced smoothly, but it does not readily predict strong coarticulatory influences of a segment C on A in an ABC sequence. Two other problems emerge below. They are that some coarticulation is difficult to characterize as anything other than overlap (for example, findings by Marchal 1988, cited above). A second is that a segment's "aggressiveness" (here, having a narrow window) in its own domain appears always to be associated with a compatible degree of aggressiveness outside of its domain, frequently beyond any transitional region between target regions.

2.2 Coproduction theories

A "coproduction" theory (Fowler, 1977) explains coarticulation as the overlapping production of—to a first approximation—invariant sequences of consonants and vowels. The context sensitivity apparent in the acoustic signal and in articulation is not "deep" context sensitivity in the sense that consonants or vowels have undergone assimilatory change (as in a feature spreading theory). Rather it is a more peripheral blending of consonants and vowels that are unchanged with respect to their essential, specified, properties.

Öhman's (1966; 1967) theory provides a seminal example of such a theory (but see also, however, (Kozhevnikov & Chistovich, 1965). In a spectrographic analysis of V_1CV_2 disyllables, Öhman noticed many instances in which the closing transitions into the consonant depended not only on V_1, but also on V_2. Likewise, transitions following consonant release depended on both vowels. X ray tracings (see also Öhman 1967) showed clear evidence that the tongue body conformation during C closure was different in the context of different flanking vowels. Öhman (1966, 166) suggested that the stop gestures were "superimposed" on a diphthongal vowel-to-vowel gesture of the tongue body and that the "tongue is able to make a distorted vowel gesture, while it is executing the stop consonant." More speculatively, he proposed three neuromuscular systems for controlling the tongue. The systems, though distinct, would use overlapping muscles. One system, the apical system, is used to produce dental, alveolar and retroflex consonants; the dorsal system produces palatal and velar consonants, and the tongue body system produces vowels. During speech production, a consonant and vowel system may be controlling the tongue in overlapping time frames and the result is "a complex summation (neural, muscular and probably mechanical also) of the responses to each of the components of the instruction." (1966, 166) Öhman's observations have been replicated many times. For example, Perkell (1969) noticed that the /k/ constriction during /həke/ consisted of a sliding movement of the tongue dorsum toward the front location for /e/. Compatible evidence of vowel-to consonant anticipatory and carryover coarticulation and sometimes vowel-to-vowel coarticulation in VCVs is provided by Barry and Kuenzel (1975), Butcher and Weiher (1976) and Carney and Moll (1971).

These findings are not captured naturally in a feature spreading account of coarticulation. The main reason is that they reveal the dynamic nature of changing articulatory parameter values during speech. Consider Perkell's finding just described. There is no change in a feature value

for /k/'s place of articulation that would yield a sliding place value. The outcome is exnlained more naturally as a growing influence of /e/'s articulatory demands during /k/.

Öhman (1967) developed a quantitative model of vowel-consonant-vowel coarticulation that did a satisfactory job of predicting the changing vocal-tract shapes (as indexed by X-ray tracings) during VCV production. Notably it includes a parameter value, k and other parameters labeled q, to implement consonant and vowel production respectively over time. To implement the temporal articulatory domain of a consonant or vowel, the associated parameter increases over time and then decreases. That is, to generate coarticulatory influences of the vowel on the consonant, for example, the vowel's influence on the vocal tract gradually waxes and then wanes. Elsewhere, we have described this waxing and waning of a segment's implementation over time as a "prominence curve" (Fowler & Smith, 1986); see Löfqvist's (1990) similar idea of "dominance").

In light of this evidence favoring coproduction, let us reconsider the data considered most supportive of feature spreading theory, evidence that lip rounding anticipates across consonant strings unspecified for rounding and that and *velum lowering anticipates across vowel strings* unspecified for nasality. Difficulties with the idea of underspecification have already been cited. More than that, however, work by Bell-Berti and her colleagues show quite convincingly that the error of accepting underspecification has led to considerable overestimation of anticipation of velum lowering and lip rounding (see also Boyce, Krakow, Bell-Berti, & Gelfer, 1990).

2.2.1 Anticipatory lowering of the velum for nasal consonants

Consider the literature on nasalization first. Researchers typically examined CV_nN strings (where Ns are nasal consonants and the subscript on the vowel signifies that different numbers of vowels intervened between C and N). Velar lowering following C was taken as evidence for onset of anticipatory nasalization from N (Moll & Daniloff, 1971). However, Bell-Berti (1980) points out that vowels are associated with lower velum heights than are oral consonants; accordingly the initial drop of the velum will be due at least to the vowel; it may or may not reflect an influence of N as well. That can be determined only by comparing CV_nN sequences with corresponding CV_nC sequences. Such a comparison indeed shows a lowering of the velum at the onset of a vowel string in CV_nC utterances that, of course, must be ascribed to the vowel rather than to coarticulatory effects of a nasal consonant (Bell-Berti & Krakow, 1991). When effects of the vowel are eliminated from velum movements in CV_nN utterances, findings are no longer consistent with feature spreading theories. Rather, *they suggest an* invariant onset of velum lowering relative to onset of nasal murmur in nasal consonant production. Bell-Berti and Harris (1981) interpret the findings as favoring a particular version of a coproduction theory, that they call "frame theory" in which the temporally-staggered onsets of component gestures of a phonetic segment are staggered in a time-invariant way.

The findings by Bell-Berti and her colleagues also help to explain an otherwise complicating finding by Bladon and Al-Bamerni (1982). Bladon and Al-Bamerni had found evidence for two patterns of anticipatory coarticulation of velum lowering—a one-step pattern of lowering, timed consistently with predictions of feature-spreading theory (that is, beginning at the onset of the first vowel in a string) and a two step pattern, the first step beginning at the onset of the first vowel and the second, as frame theory predicts, an invariant interval before the oral closing gesture for the nasal consonant. Bladon and Al-Bamerni were unable to find anything systematically different in the contexts in which each pattern was observed; therefore, they suggested that selection among the strategies was unsystematic. An alternative interpretation, however, is that sometimes the vocalic velum lowering movement (always beginning near vowel onset) overlaps completely with the lowering gesture for the nasal consonant, whereas at other times, it follows velum lowering for the vowel. Bell-Berti and Krakow (1991; see also Boyce et al., 1990) found increasing evidence of two- or multi-stage velum lowering as vocalic segments were added before the nasal consonant. Likewise, of their three talkers, one produced the target words at a considerably faster rate than the others and that subject showed a one-stage lowering pattern for all but the longest vowel segments. Finally, one talker who produced the words at two rates showed two- or multi-stage lowering only at the slower rate.

Overall, the findings on anticipatory velum lowering—originally considered to provide strong evidence in favor of a feature spreading theory of coarticulation, do not; rather, they provide better support for the view that coarticulation is coproduction. Notice, too, that Keating's targets and connections account must at least be modified

to fit the data. In particular, the model does not predict that target windows for successive segments will overlap; however, the data just described shows convincingly that they do. That is, this model too must admit the possibility of coproduction. Coarticulation is not wholly finding the most efficient pathway from one target window to another; sometimes windows overlap.

2.2.2 Lip rounding

The literature on lip rounding, like that on nasalization, has failed to support the feature spreading account. Generally, it supports frame theory. As Kent and Minifie (1977) pointed out, contradictory evidence was available even on one study commonly cited as supporting feature spreading, namely that of Benguerel and Cowan (1974). In their findings more than half the time, rounding spread not only through a preceding consonant string, but beyond it into a preconsonantal unrounded vowel. Bell-Berti and Harris (1979) obtained similar results for both of their speakers. The study by Bell-Berti and Harris (1979) and a later one (1982) showed a generally invariant relation between onset of EMG (orbicularis oris muscle of the lips) for a rounded vowel and measured acoustic onset of the rounded vowel over a variable number of prevocalic consonants.

The research by Bell-Berti and Harris tested for and found lip EMG activity for /l/, one of the consonants in the strings they used as stimuli. As noted earlier, other investigators have found rounding for other consonants. These consonantal influences on lip configuration are likely to have contaminated estimates of onset of lip rounding in the earlier research in the same way that the vocalic influences on velum height contaminated estimates of onset of velum lowering for nasal consonants. These contaminating influences can only be identified by examining control utterances that lack the specified segment (that is, VC_nV utterances in which both vowels are unrounded), and investigators have not done that generally. However, using appropriate control utterances, Boyce (1988) has shown that overlapping consonantal and vocalic lip movements approximately add so that effects of consonants on the lips in a utterance such as /kuktluk/ can be eliminated by subtracting the movement trace from /kiktlik/ from it. Whereas Boyce did not then test for the invariance of EMG onset relative to acoustic onset of the rounded vowel that Bell-Berti and Harris had reported earlier, she did find a clear intervocalic trough in lip movement activity and bimodal peaks of EMG activity in utterances with two rounded vowels. The pair of findings suggests that during the consonantal string /ktl/, rounding from the first vowel wanes while that for the second vowel increases. Hence there are two distinct rounding gestures that wax and wane in the consonantal string—just as Öhman's account of vowel-consonant production proposed. There is not a spreading of a rounding feature from vowel to consonant. Compatibly, Gelfer, Bell-Berti, and Harris (1989) super-imposed graphs of lip EMG activity (orbicularis oris) for utterances such as /ist#tu/ and /ist#ti/ having varying numbers of intervocalic consonants and final /u/ or /i/. By eliminating the activity common to both utterances, and hence due to the consonant string, they were able to identify the onset time of EMG activity associated with the rounded vowel itself. Onset times bore a near-invariant relation to release of the occlusion of the final consonant in strings of two or more consonants.

2.2.3 Lingual coarticulation

The literature on coarticulation involving the tongue supports and augments the idea of coarticulation as gestural overlap. Öhman's model suggests that demands on the articulators made by a segment increase gradually over time and decrease gradually. The serial ordering of segments in articulation is maintained not by preserving discreteness of segment production along the time axis, but, rather perhaps, by maintaining a serial ordering of their times of maximum control in the vocal tract. In addition, however, segments differ one from the other in the strengths of demands they place on different articulators (or on different articulatory systems; see below under "Coordination" and cf. Keating's idea of windows discussed above). The differences in strength have an observable consequence that is described differently (e.g., Farnetani, 1990) depending on where it is observed. If discrete domains are identified for segments in an utterance by drawing boundaries at points where coarticulating segments shift in their relative dominance in the vocal tract, then one can say that in their own domain, segments that make strong demands on an articulator "resist" coarticulatory influences from neighbors (Bladon & Al-Bamerni, 1976); in the domains of near neighbors, they exert a strong coarticulatory influence. From the perspective of a coproduction theory, resistance to coarticulation and a strong coarticulatory influence covary because they are really the same thing—namely a segment's exerting a relatively strong influence on articulators throughout its temporal domain.

Recasens (1984; 1985; 1987; in press) has conducted much of the work that has uncovered variation in coarticulation resistance in movements involving the tongue dorsum. In general, resistance to coarticulation of a consonant or vowel is associated with the amount of tongue dorsum-palatal contact associated with production of the segment (see also Farnetani, 1990). Compatibly, using acoustic and electropalatographic measures, Recasens (1984; 1987) found a decrease in vowel-to-vowel coarticulation in VCV sequences in which C is produced with considerable contact between the tongue dorsum (an important articulator in vowel production) and the palate. For example, there is less V-to-C coarticulation across palatal /j/ than across dental /n/. Compatibly, the vowel /i/, which requires a constriction in the palatal region resists consonant-vowel coarticulatory influences more so than do other vowels (Recasens, 1985), and it resists vowel-to-vowel coarticulatory overlap as well (Recasens, 1987; in press). In addition, as noted earlier, segments such as /i/ that are resistant to coarticulation in their own coarticulatory domains themselves exert strong coarticulatory influences on neighbors (see Tables II-VI in Recasens, 1987; see also Butcher & Weiher, 1976; Farnetani, Vagges, & Magno-Caldognetto, 1985).

It may be tempting to conclude from this research that production of consonants and vowels is context sensitive after all in that coarticulatory anticipation of V2 in a VCV sequence must be delayed and reduced if V1 is /i/ as compared to /a/ or if C is /j/ as compared to /n/. However, possibly, the planned segment can be invariant, while its surface manifestations vary according to its neighbor's patterns of coarticulation resistance. Consider, by analogy, the different surface consequences of an invariant squeezing action of the hand depending on whether the hand is empty, or else holding a sponge or a rock. The outcome at the surface is different both in the extent to which the hand (metaphorically, the segment being produced) closes and in the extent that it deforms the sponge (a little coarticulation resistance) and the rock (a lot of resistance). Perhaps by the same token, an invariant plan for a segment can have different surface consequences if coarticulation resistance is implemented as a real physical variable in the vocal tract. There is one striking outcome reported by Recasens (1984) that suggests exactly that. He reported instances both of anticipatory and of carryover coarticulation in which coarticulatory effects were discontinuous. That is, vowel-to-vowel effects were observed in VCV sequences even though, in consonants with considerable tongue dorsum/palatal contact, vowel-to-consonant coarticulation was absent. It is unlikely that talkers plan to begin production of V2 in V1, to stop production of V2 during C, and to recommence its production after C. An analogous plan for carryover coarticulation is even less likely.

2.3 Some tentative conclusions about coarticulation

The findings just reviewed suggest the following summary. Each consonant or vowel of the language is implemented by one or more vocal-tract actions. Actions are of two varieties: gestures (Browman & Goldstein, 1986) that are linguistically significant (and contrastive) and other, noncontrastive, ones that may occur because they are easier to produce than to suppress. Gestures for a segment may be timed or phased invariantly one with respect to another as frame theory proposes. Each vocal tract gesture has a prominence pattern of increasing then decreasing articulatory strength, where prominence refers to the extent to which the gesture exerts an influence on the character of movements in the vocal tract. Vocal tract actions differ one from the other in relative strength so that, for example, demands of /j/ or /i/ on the tongue dorsum-palate relation exceed those of /n/ and /a/. The extent to which a segment-specific action influences what is happening in the vocal tract at any point in time reflects the strength of that action and its strength relative to that of other ongoing actions affecting the same vocal-tract structures. "Strength" appears to be implemented in such a way that its effects arise at the articulatory surface, not in differential planning for a segment depending on its context. The account is incomplete in a variety of ways, lacking detail in important areas, including a specification of how strength variations are realized. It is also too simple in some respects. In particular, patterns of relative timing of gestures for a segment are not invariant—they may vary over position in a syllable as Krakow (1989) has shown for the relative timing of velum lowering and lip closing actions for syllable-initial and -final /m/. They are likely to vary over stress and rate manipulations as well. In short, the state-of-the art in coarticulation research leaves investigators still with many problems to tackle.

3. COORDINATION

From the perspective of a coarticulating segment encroaching on the domain of a second segment, the second segment applies restrictions on

where and to what extent encroachment can occur ("coarticulation resistance"). Accordingly, coarticulation by the same segment in the same (anticipatory, carryover) direction will be differentially manifested depending on the nature and strength of the restrictions applied in its coarticulatory field. Looked at from the perspective of the influenced segment, however, the restrictions are the segment's own identity; they are actions or postures the achievement of which counts as production of that segment. Somehow realization of the segment correspondingly prohibits contradictory actions. Here I examine implementation of those restrictions in speech production.

The vocal tract includes large numbers of muscles and structures that the muscles move or deform. Relative to the catalogue of movements that could occur were contractions of all possible combinations of vocal tract muscles used and contractions of all possible magnitudes, the movements that do occur in speech are limited in number and in kind. They are constrained, of course, to structure the air so that listeners can hear them. But more than that, they are low-dimensional movements—movements with order that spans groups of muscles and groups of vocal-tract structures. They are, indeed, coordinated actions.

Coordination achieves several things. Most importantly, structures of the vocal tract work together to achieve some end. For example, in production of /b/, the jaw and lips work together to achieve bilabial closure. The couplings among structures also preclude actions that violate the couplings; thereby they prohibit coarticulatory influences that would prevent the goal of the coordinative linkages. They do not completely eliminate variability or flexibility, however. For example, bilabial closure is realized with a variety of contributions by the jaw and lips. When /b/ is coarticulated with an open vowel, the jaw is lower during closure, and hence the lips do more of the closing work, than when /b/ coarticulates with /i/. Research using a perturbation procedure (e.g., Abbs & Gracco, 1984); (Kelso, Tuller, Vatikiotis-Bateson et al., 1984; Shaiman, 1989) helps to expose couplings across structures of the vocal tract. In one of these experiments, Kelso, Tuller, Vatikiotis-Bateson, and Fowler (1984) asked talkers to produce "It's a ___ again," with /baeb/ or /baez/ serving as target syllable. On a low proportion of trials, randomly selected, during the closing gesture for the second /b/ in /baeb/ or for the /z/ in /baez/, the talker's jaw was unexpectedly braked, preventing its normal contribution to closure for the consonantal constriction. On perturbed relative to unperturbed trials, within 20-30 ms of the perturbation in /baeb/, the orbicularis oris muscle of the upper lip showed extra activation and by achievement of closure, the lip had moved farther down than on unperturbed trials. If the jaw was braked during closing for /z/, extra activation was observed in the genioglossus muscle of the tongue allowing the tongue to compensate for the unusually low position of the jaw. The upper lip did not show the same extra downward movement on /z/-perturbed trials that it showed on /b/-perturbed trials. Other research (Shaiman, 1989) shows that when an articulator of the vocal tract is perturbed that is not involved in a consonantal closing gesture, closing on perturbed and unperturbed trials is alike. In short, the responses to perturbation are adaptive and they reveal a coupling among selective articulators of the vocal tract that jointly achieve some phonetic gestural end. Coupled structures and their neuromuscular underpinnings are know as "synergies" or "coordinative structures." Whereas Löfqvist (1990) suggests that there are no dynamic perturbations in speech analogous to a jaw pull, perhaps there are. Coarticulatory encroachments from low vowels can perturb a talker's jaw, pulling it down during closure for /b/. Possibly, then, the couplings serve two functions; they bring about the coordinated action that constitutes a linguistic gesture of the vocal tract, and they permit only those coarticulatory encroachments that will not prevent the gesture from being realized.

The short-latency responses to the perturbations suggest that the couplings are low- level. That is, they are not cognitive couplings, but, rather neuromuscular ones. This may help to rationalize findings by Recasens summarized earlier of discontinuities in coarticulatory influences. Whereas it would be surprising for speakers to plan for V-to-V coarticulatory influences, yet plan for no V-to-C influences in a VCV sequence, the finding of discontinuities in coarticulation is less surprising if segments are planned to have an invariant coarticulatory field that then gets differentially suppressed by other synergies active in the vocal tract.

Following Browman and Goldstein (1986; 1989), we may call the vocal tract actions of a synergy a "phonetic gesture" or, more simply, a "gesture." Phonetic gestures are, then, linguistically significant actions of the vocal tract. In the research using the perturbation technique just described, perturbations disrupted movements by one articulator among two or more that participated in a phonetic gesture. That is, perturbations and compen-

sations were interarticulatory, but intragestural. However, some phonetic segments are defined by more than one gesture, and the timing or phasing between or among gestures may also be crucial to the identity of the segment. For example, the timing of an oral constriction gesture and a glottal devoicing gesture determines whether a consonant is preaspirated or aspirated (see, e.g., Löfqvist, 1980); Löfqvist & Yoshioka, 1984). Presumably, then, intrasegmental gestures must be coupled and one should see evidence of the coupling in perturbation experiments. To date, there is little evidence on the topic.

However, Munhall, Löfqvist, and Kelso (1988) have perturbed the lower lip during closing from a vowel to a /p/. The perturbation delayed achievement of closure, thereby lengthening the vowel. However, onset of glottal opening for /p/ was also delayed, giving rise to a perceptually adequate aspirated /p/. (Even so, there was disruption of the coordinative relation between the gestures such that the voice-onset times on perturbed trials were unusually long.)

Another index, perhaps, of a coupling relation between the gestures of a segment is provided by tests for invariant relative timing (as summarized in Löfqvist 1990). Coupling between gestures of a segment should give rise to invariance of relative timing between the gestures so that, as the segment is produced at various rates or with different levels of stress, temporal intervals between gesture onsets scale proportionately to changes in other intervals produced by the coupled actions. (The idea is that if the gestures are products of a common synergy, and rate changes are achieved by changes in a parameter that is common to the synergy, all temporal intervals produced by the gestures will scale proportionately.) Löfqvist (1990) applied a test for proportionality of intervals proposed by Gentner (1987) to several sets of data including measures of intrasegmental- intergestural intervals and intersegmental- intergestural intervals. Whereas 90% of tests for proportional changes in intervals over variation in rate and stress were rejected in tests of the latter intervals, just 33% were rejected in tests of the former intervals. Löfqvist does not consider this particularly strong support for the proportional-durational test of coupling between gestures of a segment, because the reason why 67% of tests failed to reject the hypothesis of proportional durations for intrasegmental-intergestural intervals was not that intervals were relatively invariant, but rather because they were extremely noisy (see his Figures 11-15). Even so, his data do reveal marked differences in the temporal relations among gesture belonging to the same and to different phonological segments, with the latter relations showing systematic departures from the proportional-duration hypothesis and the former showing only unsystematic departures.

4. SPEECH DYNAMICS

There is a new development in the study of speech production that I will describe only briefly. It is as yet relatively untried; however, it promises to have a marked influence on research in the field. Although speech production is remarkable as a motor activity, it is not wholly unique. Some common issues arise in investigations of a variety of intentional motor skills. More fundamentally, however, some theorists suggest that intentional actions in general (Kugler & Turvey, 1987); Kugler, Kelso, & Turvey 1980) and speech production in particular (Saltzman, 1986; Saltzman & Kelso 1987; Saltzman et al., 1989; Kelso & Tuller, 1984) constitute a special instance of "self-organization" in physical systems. Accordingly, they may be best understood by embedding their investigation in the larger context of the study of self-organizing physical systems. Complex physical systems that are open to the flow of energy from the environment, whether they are living systems or not, develop macroscopic, low dimensional patterned and stable activities that can be modeled as attractors of just a few sorts. Most simply, a physical system can be modeled as a "point attractor" if, when perturbed, it tends to return to the same final target—much as the vocal tract does if it is perturbed during bilabial closure (e.g., Saltzman & Kelso, 1987).

Saltzman and colleagues have shown that many central features of speech production—including adaptive responses to perturbations and consequences of coarticulatory overlap (see Saltzman & Munhall, 1989) can be modeled if phonetic gestures are modeled as dynamical systems. On the other side, Tuller and Kelso (1990) have shown that speech production exhibits some of the central characteristic features of dynamical systems. Finally, Browman and Goldstein (1986) have developed an "articulatory phonology" whose primitive units, phonetic gestures, are defined by dynamical parameters of the vocal-tract point attractors of Saltzman's articulatory ("task-dynamic") model. Possibly, embedding the investigation of speech production in the context of studies of complex open physical systems generally will help to deepen our understanding of synergies and

their achievement of low-dimensional, coordinated actions. In turn, understanding of these physical systems may literally add substance to the linguist's concepts of phonological segments and their features.

REFERENCES

Abbs, J., & Gracco, V. (1984). Control of complex gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology, 51,* 705-723.

Barry, W., & Kuenzel, H. (1975). Co-articulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics, 3,* 263-282.

Bell-Berti, F. (1980). Velopharyngeal function: A spatio-temporal model. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 291-316). New York: Academic Press.

Bell-Berti, F., & Harris, K. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America, 65,* 1268-1270.

Bell-Berti, F., & Harris, K. (1981). A temporal model of speech production. *Phonetica, 38,* 9-20.

Bell-Berti, F., & Harris, K. (1982). Temporal patterns of coarticulation: lip rounding. *Journal of the Acoustical Society of America, 71,* 449-454.

Bell-Berti, F., & Krakow, R. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America., 90,* 112-123.

Benguerel, A., & Cowan, H. (1974). Coarticulation of upper lip protrusion in French. *Phonetica, 30,* 41-55.

Bladon, A., & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics, 4,* 137-150.

Bladon, A., & Al-Bamerni, A. (1982). One-stage and two-stage temporal patterns of coarticulation. *Journal of the Acoustical Society of America, 72,* S104.

Boyce, S. (1988). *The influence of phonological structure on articulatory organization in Turkish and English: Vowel harmony and coarticulation,* doctoral dissertation, Yale University.

Boyce, S., Krakow, R., Bell-Berti, F., & Gelfer, C. (1990). Converging sources of evidence for dissecting articulatory movements into gestures. *Journal of Phonetics, 18,* 173-188.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan, & J. Anderson (Ed.), *Phonology yearbook* 3 (pp. 219-252). Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201-252.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics, 4,* 59-74.

Carney, P., & Moll, K. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica, 23,* 193-202.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics, 1,* 239-248.

Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research, 11,* 707-721.

Delattre, P., & Freeman, D. (1968). A dialect study of American rs by X-ray motion picture. *Linguistics, 44,* 29-68.

Fant, G., & Lindblom, B. (1961). Studies of minimal speech and sound units. *Speech Transmission Laboratory: Quarterly Progress Report, 2,* 1-21.

Farnetani, E. (1990). V-C-V lingual coarticulation and its spatiotemporal domain. In W. J. Hardcastle, & A. Marchal (Ed.), *Speech production and speech modeling* (pp. 93-130). The Netherlands: Kluwer.

Farnetani, E., Vagges, K., & Magno-Caldognetto, E. (1985). Coarticulation in Italian /VtV/ sequences: A palatographic study. *Phonetica, 42,* 78-99.

Fowler, C., Rubin, P., Remez, R., & Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production I: Speech and talk* (pp. 373-420). London: Academic Press.

Fowler, C., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Ed.), *Invariance and variability of speech processes* (pp. 123-136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, C. A. (1977). *Timing control in speech production.* Bloomington: Indiana University Linguistics Club.

Fowler, C. A. (1990). Comments on the contributions by Pierrehumbert and Nearey. *Journal of Phonetics, 18,* 425-434.

Gelfer, C., Bell-Berti, F., & Harris, K. (1989). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America, 86,* 2443-2445.

Gentner, D. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review, 94,* 255-276.

Henke, W. (1966). *Dynamic articulatory models of speech production using computer simulation,* doctoral dissertation, MIT.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). Orlando: Academic Press.

Keating, P. (1988a). Mechanisms of coarticulation: Articulatory evidence. *UCLA Working Papers in Phonetics, 69,* 3-29.

Keating, P. (1988b). Underspecification in phonetics. *Phonology, 5,* 275-292.

Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamic principles for speech and movement control. *American Journal of Physiology, 246,* R982-R935.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 812-832.

Kent, R., & Minifie, F. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5,* 115-144.

Klatt, D. (1973). Vowel-lengthening is syntactically determined. *Journal of Phonetics, 3,* 129-140.

Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and perception* (30, 543). Washington, DC: Joint Publications Research Service.

Krakow, R. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures,* doctoral dissertation, Yale University.

Kugler, P., Kelso, J. A. S., & Turvey, M. (1980). On the concept of coordinative structures as dissipative structures, I. Theoretical lines of convergence. In G. Stelmach & J. Requin (Ed.), *Tutorials in motor behavior* (pp. 3-47). Amsterdam: North Holland.

Kugler, P. & Turvey, M. (1987). *Information, natural law and the self-assembly of rhythmic movement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Leidner, D. (1973). *An electromyographic and acoustic study of American English liquids,* doctoral dissertation, University of Connecticut.

Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics, 8,* 475-490.

Löfqvist, A. (1990). Speech as audible gestures. In W. J. Hardcastle & A. Marchal (Ed.), *Speech production and speech modeling* (pp. 289-332). The Netherlands: Kluwer.

Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in vowel-consonant production. *Speech Communication, 3,* 279-289.

Magen, H. S. (1989). An acoustic study of vowel-to-vowel coarticulation in English, doctoral dissertation, Yale University.

Manuel, S. (1987). *Acoustical and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages*, doctoral dissertation, Yale University.

Manuel, S., & Krakow, R. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Reports on Speech Research, SR-77/78*, 69-87.

Marchal, A. (1988). Coproduction: Evidence from EPG data. *Speech Communication*, 7, 287-295.

Moll, K. (1962). Velopharyngeal closure in vowels. *Journal of Speech and Hearing Research*, 36, 30-77.

Moll, K., & Daniloff, R. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, 50, 678-684.

Munhall, K., Löfqvist, A., & Kelso, J. A. S. (1988). Lip-larynx coordination in speech: A perturbation study. *Journal of the Acoustical Society of America*, 80, S109.

Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39, 151-168.

Öhman, S. (1967). Peripheral motor commands in labial articulation. *Speech Transmission Laboratory: Quarterly Progress Report*, 4, 30-63. Perkell, J. (1969). Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, MA: MIT Press.

Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei. *Journal of the Acoustical Society of America*, 32, 693-703.

Pierrehumbert, J. (1990). Phonological and phonetic representations. *Journal of Phonetics*, 18, 375-394.

Recasens, D. (1984). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics*, 12, 61-73.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulation resistance in Catalan CV sequences. *Language and Speech*, 28, 97-114.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, 15, 299-312.

Recasens, D. (1989). Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8, 293-308.

Repp, B. (1981). On levels of description is speech research. *Journal of the Acoustical Society of America*, 69, 1462-1464.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators. In H. Heuer, & C. Fromm (Ed.), *Generation and modeling of action patterns*. New York: Springer-Verlag.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled action: A task-dynamic approach. *Psychological Review*, 94, 84-106.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America*, 86, 78-88.

Tuller, B., & Kelso, J. A. S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 429452) Hillsdale, NJ: Lawrence Erlbaum Associates.

FOOTNOTES

*In G. Blanken, J. Dittman, H. Grimm, J. Marshall, & C-W Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook*. Berlin: Walter de Gruyter (in press).

† Also Dartmouth College.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 13-26

Characteristics of Speech as a Motor Control System*

Vincent L. Gracco

The structural and functional organization of any biophysical system provides potentially important information on the underlying control structure. For speech, the anatomical and physiological components of the vocal tract and the apparent functional nature of speech motor actions suggest a characteristic control structure in which the entire vocal tract can be viewed as the smallest functional unit. Sounds are coded as different relative vocal tract configurations generated from neuromuscular specifications of characteristic articulatory actions. Sensorimotor processes are applied to the entire vocal tract to scale and sequence changes in vocal tract states. Sensorimotor mechanisms are viewed as a means to predictively adjust speech motor output in the face of continuously changing peripheral conditions. An underlying oscillatory process is hypothesized as the basis for sequential speech movement adjustments in which a centrally-generated rhythm is modulated according to internal (task) requirements and the constantly changing configurational state of the vocal tract.

Speaking is a complex action involving a number of levels of organization and representative processes. At a cognitive level, speaking represents the manipulation of abstract symbols through a synthesis of associative processes expressed through a sophisticated linguistic structure. At a neuromotor level, at least seven articulatory subsystems can be identified (respiratory, laryngeal, pharyngeal, lingual, velar, mandibular, and labial) which interact to produce coordinated kinematic patterns within a complex and dynamic biomechanical environment. At an acoustic level, characteristic patterns result from complex aerodynamic manipulations of the vocal tract. The cognitive, sensorimotor and acoustic processes of speech and their interactions are critical components to understanding this uniquely human behavior. As the interface between the nervous system and the acoustic medium for speech production/perception, speech motor processes constitute a direct link between higher level neurophysiological processes and the resulting aerodynamic/acoustic events.

The author thanks E. V.-Bateson and C. Fowler for editorial comments and to Y. Manning-Jones for word processing. The writing of this paper was supported by NIH grants DC-00121 and DC-00594.

In the following chapter, characteristics of the speech motor control process will be evaluated from a functional perspective emphasizing the structural and functional organization of the vocal tract and the timing characteristics associated with their continuous modulation. In contrast to perspectives which emphasize the large numbers of muscular/kinematic degrees of freedom, the current perspective is one that assumes that the overall vocal tract is the smallest unit of functional behavior. Sounds are encoded according to characteristic vocal tract shapes specified neuromuscularly and modulated through sensorimotor mechanisms to adapt to the constantly changing peripheral environment. Examination of the structural components and their interaction is consistent with this macroscopic organization as are a number of empirical observations. The functional organization is implemented by a limited number of sensorimotor control processes that scale overall vocal tract actions spatiotemporally within a frequency-modulated rhythmic organization characteristic of more automatic, innate motor behaviors.

Structural Properties

In order to describe speech from the perspective of a motor control system, a necessary step is to identify the components of the motor system to

determine how their structural properties may reflect on the overall functional organization. The structures of the vocal tract include the lungs, larynx, pharynx, tongue, lips, jaw, and velum. Anatomically the vocal tract structures display unique muscular architecture, muscular connections, and muscular orientation that determine their potential contributions to the speech production process. For example, the orientation of the muscles of the pharynx, primarily the pharyngeal constrictors, is such that they generate a sphincteric action on the long axis of the vocal tract producing a change in the cross-sectional area and the tension or compliance of the pharyngeal tissues. The muscles of the velum are oriented primarily to raise and lower the soft palate separating the oral and nasal cavities. Perioral muscles are arranged such that various synergistic muscle actions result in a number of characteristic movements such as opening and closing of the oral cavity and protruding and retracting the lips. Some of the components, such as the tongue and larynx, can be subdivided into extrinsic and intrinsic portions each of which appear to be involved in different functional actions. Intrinsic tongue muscle fibers are oriented to allow fine grooving of the longitudinal axis of the tongue and tongue tip and lateral adjustments characteristic of liquid and continuant sounds. Extrinsic tongue muscles are arranged predominantly to allow shaping of the tongue mass as well as elevation, depression and retraction of portions of the tongue. Intrinsic laryngeal muscles are arranged to open and close the glottis reciprocally and adjust the tension of the vibrating vocal folds, while extrinsic laryngeal muscles are oriented to displace the entire laryngeal complex (thyroid cartilage and associated intrinsic muscles and ligaments). Generally, movements of the vocal tract can be classified into two major categories; those that produce and release constrictions (valving) and those that modulate the shape or geometry of the vocal tract. The valving and shaping actions are generally associated with the production of consonant and vowels sounds, respectively (Öhman, 1966; Perkell, 1969).

In addition to the structural arrangement of the vocal tract muscles for valving and shaping actions, mechanical properties of individual vocal tract structures provide insight into the functional organization of the speech motor control system. The dynamic nature of the tissue load against which the different vocal tract muscles contract is extremely heterogeneous. For some structures such as the lips and vocal folds, inertial considerations are minimal, while for the jaw and respiratory structures inertia is a significant consideration. The tongue and lips are soft tissue structures that undergo substantial viscoelastic deformation during speech while the jaw and perhaps the lips display a degree of anisotropic tension (Lynn & Yemm, 1971). Even seemingly homogeneous structures such as the upper and lower lips, display different stiffness properties (Ho, Azar, Weinstein, & Bowley, 1982) possibly contributing to their differential movement patterns (Gracco & Abbs, 1986; Gracco, 1988; Kelso et al., 1984). Considering the structural arrangement of the vocal tract, the different muscular orientations and the vast interconnection of muscles, cartilages, and ligaments it is clear that complex biomechanical interactions among structures are the rule. Passive or reactive changes in the vocal tract due to inherent mechanical coupling is a consequence of almost any vocal tract action, with the relative significance varying according to the specific structural components and conformational change and the speed at which adjustments occur. As a result, a single articulatory action may generate primary as well as secondary effects throughout the vocal tract. The examination of individual articulatory actions are important to determine their contribution to the sound producing process. However, individual articulatory actions never have isolated effects. The combination of the viscoelastic properties of the tissues, the different biomechanical properties of vocal tract structures, and the complex geometry of the vocal tract comprise a complex biomechanical environment. The kinematic and acoustic variability characteristic of speech production reflects in part the differential filtering of neural control signals by the peripheral biomechanics. Only through detailed biophysical models of the vocal tract and considerations of potential biomechanical interaction associated with various phonetic environments can the control principles of the speech motor control system be separated from structural or cognitive/linguistic influences.

Functional Organization

In order to characterize the speech motor control system accurately, and pose the motor control problem correctly, it is important to determine how the behavior is being regulated. That is, are the individual sound-influencing elements being independently controlled or does the control structure involve larger units of behavior, and if so, what is the organizational structure? For speech, the simple observation that even an isolated vowel

sound requires activity in respiratory muscles, tension and adduction of the vocal folds adjustments in the compliance of the oropharyngeal walls, shaping of the tongue, positioning of the jaw, elevation of the velum, and some lip configuration is rather convincing evidence that speech is functionally organized at a level reflecting the overall state of the vocal tract. It is the interaction of all the neuromuscular components that provide each speech sound with its distinct character, not the action of any single component. The often-cited fact that speech production involves over 70 different muscular degrees of freedom, while perhaps anatomically factual, is a functional misrepresentation of the motor control system organization. As early as the birth cry and through the earliest stages of speech development, the infants vocalizations involve the cooperative action of respiratory, laryngeal, and supralaryngeal muscles to produce sounds. A similar observation can be made for locomotion in that rhythmic stepping and other seemingly functional locomotion-like behaviors can be elicited well before the infant manifests upright walking (Thelen, 1985, 1986). It appears that functional characteristics of many human behaviors are present at birth or very early in the infants development suggesting that the "significant functional units of action" (Greene, 1972) may be innate properties of the nervous system. It is suggested that speech motor development reflects the ability to make finer and more varied adjustments of the vocal tract, not the mastering of the articulatory or muscular degrees of freedom.

As suggested above, the characteristics of speech as a motor control system include a control structure in which the smallest functional unit is the entire vocal tract. Recent studies have demonstrated examples of large scale manipulation of vocal tract actions rather than the modulation of separate articulatory actions. As shown in Figure 1, movements of individual articulators such as the upper lip, lower lip, and jaw demonstrate timing relations such that adjustments in one structure are accompanied by adjustments in all functionally-related structures.

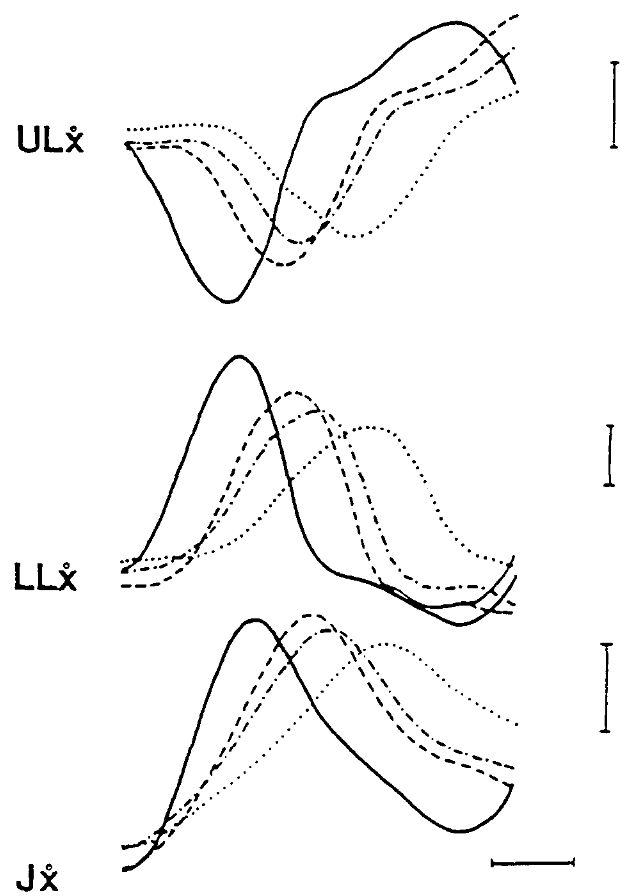

Figure 1. **Upper Lip (UL), Lower Lip (LL), and Jaw (J) movement velocities associated with the first "p" closing in "sapapple." As the preceding vowel duration changes, the timing of the UL, LL, and J change in a consistent and unitary manner (from Gracco, 1988). Calibration bars are 50 mm/sec (vertical) and 100 ms (horizontal).**

The coordinative process reflects a constraint on articulatory actions involved in the production of a specific sound. Similar results can be observed for other more spatially remote, but functionally related articulators. As shown in Figure 2, movements of the larynx and the lower lip demonstrate a similar timing dependency for the production of the "f" in "safety". In order to generate the frication noise characteristic of the /f/, the glottal opening and labial constriction is appropriately timed. As the timing of one structure changes, the timing of the other functionally-related articulatory action also changes. Similarly, for movements associated with resonance producing vowel events, timing constraints can be observed between laryngeal voicing and jaw opening associated with tongue positioning for a vowel (Figure 3). Here, the laryngeal action associated with phonation and the change in jaw positioning to assist the tongue in vowel production demonstrate similar coordinative interdependency. Some preliminary evidence further suggests that certain physiological changes associated with the production of emphatic stress results in an increase in the actions of all portions of the vocal tract rather than being focused on one specific articulator (Fowler, Gracco, & V.-Bateson, 1989). In the presence of a potentially disruptive mechanical disturbance applied to one of the contributing articulators there is a tendency for the timing of all articulators to readjust (Gracco & Abbs, 1988). The timing of individual articulators is apparently not adjusted singularly but reflects a system level organization (see Löfqvist & Yoshioka, 1981; 1984; Tuller, Kelso, & Harris, 1982; for other examples). It is not clear how general these observation are with regard to all speech sounds in all possible contexts. For example, the lip/jaw and laryngeal/supralaryngeal coordination observed in Figures 1 and 2 is modified when the sound is at the beginning of a word apparently reflecting a change in the functional requirements of the task. The importance of these kinds of observations is not the specific observable pattern but the presence of characteristic patterns that are used for time-dependent articulatory adjustments.

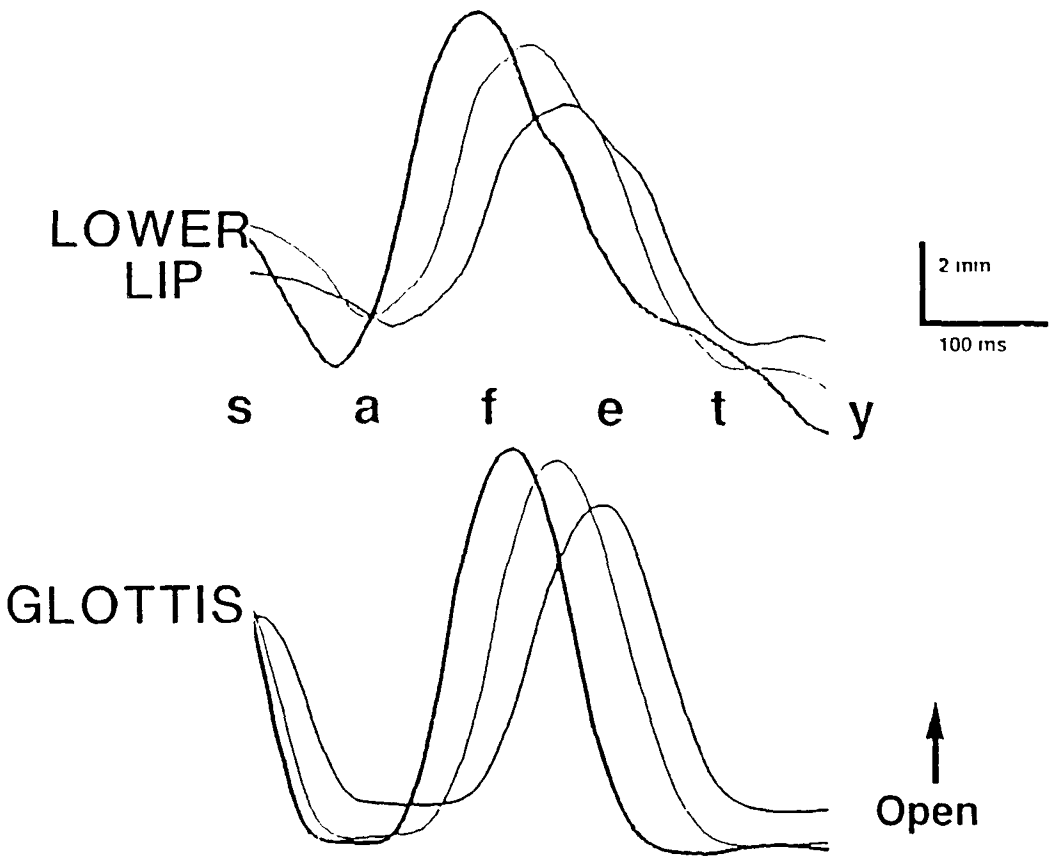

Figure 2. **Lower Lip closing and glottal opening movements for three repetitions of the word "safety." As the lower lip closing movement for "f" varies, the timing of the glottal opening (devoicing) also varies (from Gracco & Löfqvist, 1989). Similar to Figure 1, the timing of the oral and laryngeal actions appear to be adjusted as a unit.**

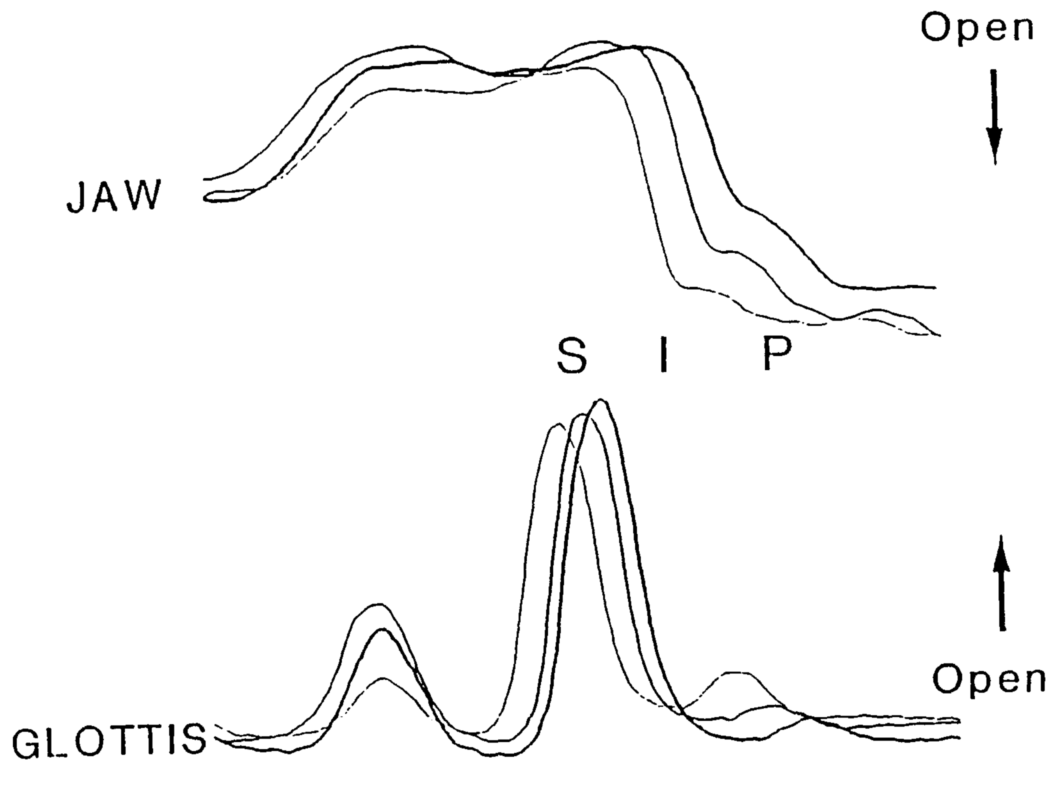

Figure 3. **Timing relations between the glottal closing and the jaw opening associated with the vowel in "sip." As the glottal opening/closing associated with the "s" and subsequent vowel varies, the jaw opening (noted by the downward movement) also varies proportionally (from Gracco & Löfqvist, 1989).**

Speech motor patterns reflect characteristic ways of manipulating the vocal tract, in the presence of a constant pressure source, to generate recognizable and language-specific acoustic signals (Ohala, 1983). The process through which such functional cooperation occurs has been described for many motor tasks in various contexts with the assumption that the control actions involve the *assembly* of functional units of the system organized into a larger systems known as synergies or coordinative structures (Bernstein, 1967; Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980; Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Saltzman, 1979; 1986; Turvey, 1977; Kugler, Kelso, & Turvey, 1980; 1982). In keeping with the interactive structural configuration outlined previously and the apparent functional nature of the task itself, a modification of this perspective is offered. Speaking appears to involve coordinative structures (or synonymously motor programs; see Abbs, Gracco, & Cole, 1984; Gracco, 1987) available for all characteristic vocal tract actions associated with the sound inventory of the language. It is not the case, however, that a coordinative structure or a motor program is a process but a set of sensorimotor specifications identifying the relative contribution of the vocal tract structures to the overall vocal tract configuration (see Abbs et al., 1984; Gracco, 1987). As such, coordinative structures may be more rigidly-specified than previously thought and the distinction between a flexible coordinative structure and a hard-wired motor program algorithm may be more rhetorical than real (cf. Kelso, 1986 for discussion of differences). In this regard, two observations are of note. When the contribution of jaw movement is eliminated, by placing a block between the teeth, jaw closing muscle actions are still present (Folkins & Zimmermann, 1981). Further, in response to jaw perturbation, both functionally-specific responses and non-functional responses are observed such as upper lip muscle increases when the subjects are not producing sounds requiring upper lip move-

ment (Kelso Tuller, V.-Bateson, & Fowler, 1984; Shaiman, 1989). Together, these observations reflect on specific aspects of the speech motor control process and suggest that speech production may rely to some degree on fixed neuromuscular specifications. The presence of jaw muscle actions when the jaw movement is eliminated is consistent with the previous suggestion that speech motor control is a wholistic process involving the entire vocal tract. The presence of upper lip muscle increases (albeit small) when the sound being produced does not involve the upper lip, reflects on the underlying control process. The interaction of the phasic stimulus (from the perturbation) with activated motoneurons will produce the functionally-specific compensatory response. If the motoneurons are inactive, or slightly active, the phasic stimuli would result in small increases in muscle activation levels without any significant movement changes. This is a much simpler control scheme in that certain interactions and functionally-specific responses are a consequence of the activation of specific muscles and the actual synaptic interactions of various vocal tract structures (Gracco, 1987). The advantage of this perspective is that certain properties of speech production *result from the physiological organization* and focus the functional organization of the speech motor control system on the neural coding of speech sounds and the characteristic sensorimotor processes that modulate and sequence vocal tract configurations.

Neural Coding of Speech Motor Actions

The coding of speech is viewed as the process by which overall vocal tract states are "represented and transformed by the nervous system" (see Perkel & Bullock, 1968). This coding is similar to what has previously been identified as the selection of muscular components associated with a specific motor act (cf. Evarts, Bizzi, Burke, DeLong, & Thach, 1972. In the following, the selection of characteristic vocal tract states will be evaluated with respect to two components of the hypothetical specification process although the actual neural coding is viewed as a single process and is only presented separately for the purpose of clarity. As stated previously, the actions of the vocal tract are designed to either valve the air stream for different consonant sounds or to shape the geometry of the vocal tract for different vowel and vowel-like sounds. Considering the place of articulation for vowels and consonants naturally results in categorical distinctions which are apparent acoustically and aerodynamically (Stevens, 1972). However, rather than dichotomizing these apparently discrepant processes, it is suggested that valving and shaping can be conceptualized as a single physiological process. That is, speech sounds are coded according to overall vocal tract states which include primary articulatory synergies. When the appropriate muscles are activated, the resulting force vectors create characteristic actions resulting in vocal tract states which act to valve the pressure or change the geometry without creating turbulence producing constrictions. It is the orientation of the activated muscle fibers, the activation of synergistic and antagonistic muscles, and the fixed boundaries of the vocal tract (the immobile maxilla) that result in the achievement of characteristic shapes or constriction locations; certain muscular synergies can only result in certain vocal tract configurations. For example, selection of certain upper and lower lip muscles (orbicularis oris inferior and superior, depressor anguli oris, mentalis, depressor labii inferior) will always result in the approximation of the upper and lower lips for "p", "b", or "m". The magnitude or timing of the individual muscle actions may vary, but bilabial closure will always involve the activation of upper and lower lip muscles; otherwise bilabial closure could not be attained. Similarly, changing the focus of neural activation to regions representing lower lip muscles (orbicularis oris inferior and mentalis with primary focus in mentalis) results in movements consistent with labiodental constriction for "f" and "v" achieved against the immobile maxillary incisors (Folkins, 1976). Different relative contributions of extrinsic and intrinsic tongue muscles result in various shapes and movements on the tongue tip, blade and body resulting in characteristic constrictions or shapes as a consequence. Constriction location and constriction degree are useful categories to describe different speech sounds because they specify what is distinctive to each phonetic segment. Control over the vocal tract configuration through the development of finer control over the neuromuscular organization provides a more reasonable description of the speech acquisition process because the entire vocal tract is manipulated not just the distinctive attributes for each sound. The neuromotor differences in consonant and vowel sounds appear to be reflected in other characteristics of the control process.

One such characteristic involves the compliant states of the vocal tract consistent with the level of tension in the tissue walls. The importance of

tissue compliance can be inferred from a number of observations. A major physical difference between voiced and voiceless consonants is in the level of air pressure associated with their production. Voiceless sounds are generally produced with higher vocal tract pressures than their voiced counterparts. The pressure difference, which has significant aerodynamic and acoustic consequences, results from changes in the tension in the pharyngeal and oral cavities as well as from pressure from the lungs (Müller & Brown, 1980). For example, subjects engaged in producing speech while simultaneously engaged in a valsalva maneuver (forceful closing the glottis thereby eliminating the lung contribution) were able to maintain voiced/voiceless intraoral pressure differences apparently resulting from changes in the overall compliance of the vocal tract walls (Brown & McGlone, 1979). Together with experimental evidence that kinematic and electromyographic characteristics of lip and jaw movements are insufficient to differentiate voiced and voiceless sounds (Lubker & Parris, 1970; Harris, Lysaught, & Schvey, 1965; Fromkin, 1966), it appears that a major factor in generating voicing and voicelessness is the specification of overall vocal tract compliance. Two possible compliant states of the vocal tract are sufficient to categorize most speech sounds; low compliance associated with voiceless consonants and high compliance associated with voiced consonants and vowels. Compliant states of the vocal tract are associated with gross changes in the activity of at least the pharyngeal constrictors as has been observed (Minifie, Abbs, Tarlow & Kwaterski, 1974; Perlman, Luschei, & DuMond, 1989) and possibly other portions of the walls of the vocal tract (intraoral cavity). The specification of low compliance (resulting in high vocal tract pressures) would be associated with increased activity in larygneal muscles to assist in the devoicing gesture, and high compliance (resulting in low vocal tract pressures) would be associated with a relaxation of the muscle activity in the pharyngeal and oral cavities to allow cavity expansion for voiced stops and continuants (Bell-Berti & Hirose, 1973; Westbury, 1983; Perkell, 1969). Certain tense vowels may result from an intermediate level of compliance (between high and low) such that voicing is maintained but overall compliance is slightly higher than for lax vowels. It is important to note that modification in compliance is a process that produces a relatively slow change in the state of the vocal tract, with relaxation (high compliance) a slower process than constriction (low compliance). Together, specification of the compliant state of the vocal tract and selection of specific muscular actions is one means by which the vocal tract states may be neurally specified.

It should be noted, however, that the coding of speech motor actions is viewed primarily as a static process in which characteristic states of the vocal tract are identified prior to their actual implementation. Considering some dynamic properties of the speech motor control system provide some insight into the manner in which different sounds may acquire further acoustic and kinematic distinction. For example, lip closing movement associated with the voiceless bilabial stop "p" is generally but not consistently associated with a higher velocity than the voiced bilabial "b" or "m" (Chen, 1970; Gracco, submitted; Summers, 1987; Sussman, MacNeilage, & Hanson, 1973). Lip and jaw closing movements are initiated earlier relative to vowel onset for voiceless "p" than for voiced "b" or "m" (Gracco, submitted) resulting in shorter vowel durations. One possible explanation is that voiceless sounds are produced at a higher rate or frequency than their voiced counterparts reflecting a different underlying frequency specification. Movement frequency is one dimension along which different speech sounds can be generally categorized. This hypothetical frequency modulation can be integrated with another dynamic property of the control system. Not only are closing movements generally faster for a voiceless than for a voiced consonant, but the preceding opening movement has also been observed to be faster (Gracco, submitted; Summers, 1987). It appears that not only may sounds be coded as a function of the frequency of individual vocal tract adjustments but that the functional requirements for specific sounds may be distributed across movement cycles rather than focused on a single movement phase. This observation suggests the operation of a look-ahead mechanism (Henke, 1966) similar to or identical with the mechanism underlying anticipatory coarticulation which predictively adjusts vocal tract actions. Speech motor control is a dynamic neuromotor process in which overall vocal tract compliance, the location of primary valving or shaping synergies, and frequency-modulated motor commands are specified by the immediate and future acoustic/aerodynamic requirements.

Invariance, Redundancy, and Precision

Before presenting some of the specific processes of the speech motor control system that are used

to modulate overall vocal tract organization, two important and related issues should be addressed; invariance and precision. The search for invariance has a long and generally unsuccessful history in investigations of speech production with the obvious conclusion that invariance is not a directly observable event (alternatively, the appropriate metric has not been identified). From the perspective of speech as a motor control system, a more fundamental issue is the precision with which any quantity, variable, or vocal tract configuration is regulated. The presence of substantial acoustic, kinematic, electromyographic, and aerodynamic variability suggests that the speech motor control process operates at less than maximal precision (or within rather broad tolerance limits). The achievement of characteristic vocal tract configurations or individual articulatory actions is accomplished by a synthesis of general activation of most vocal tract structures (setting of overall vocal tract compliance) and focused activation of the relevant muscular synergies. This is consistent with neurophysiological evidence demonstrated in the studies of Kots (Kots, 1975) in which voluntary movement is seen as a synthesis of diffuse excitation (pretuning), a more fixed and discrete *increase in motoneuron excitability* (tuning) and the final "triggering" process. Similarly, brain potentials prior to the onset of muscle activity display rather diffuse activation over multiple cortical areas for discrete, finger and toe movements (Boschert, Hink, & Deecke, 1983; Deecke, Scheid, & Kornhuber, 1969) and involve larger regions for production of speech. (Curry, Peters, & Weinberg, 1978; Larsen, Skinhøj, & Lassen, 1978). One plausible perspective is that the nervous system modulates the focus of primary activation but that this process is not punctate. That is, activation and deactivation of cortical and perhaps subcortical cells involve diffuse and slow changes in activation or deactivation which result in distributed tonic and phasic muscle activity. Specification of vocal tract configurations for specific sounds may involve characteristic patterns of activation and inhibition in all vocal tract muscles with only slightly greater focus on critical articulators involved in the more dominant or sound-critical movements. In some cases muscles may be partially activated just because of the proximity of their motoneurons to other activated motoneurons. One conclusion is that the neural processes underlying speech motor control are broadly specified and that the functional speech production goals (and the requisite perceptual properties) are only categorically invariant. As suggested by the apparent quantal nature of speech (Stevens, 1972), as long as the articulatory patterns are within a certain range (have not made a category change), the corresponding phonetic properties will be perceived, with kinematic variations producing very little perceptual effect. Perhaps speech perception and production should be appropriately represented as stochastic processes based on probability statements implemented through an adequate but imprecise control system. Strict determinism, invariance, and precision are most likely relegated to man-made machines working under rigid tolerance limits or simplified specifications, not to complex biological systems.

Sensorimotor Control Processes

Similar to the temporal organization for speech, spatial interactions are evident that reflect multiarticulate manipulations to achieve characteristic vocal tract states. The clearest examples of cooperative and functionally-relevant spatial interactions are observed when one articulator, such as the lip or jaw, is disturbed during speaking. Following the application of a dynamic perturbation impeding the articulatory movement, a compensatory adjustment is observed in the articulator being perturbed as well as other functionally-related, spatially-distant articulators (Abbs & Gracco, 1984; Folkins & Abbs, 1975; Gracco & Abbs, 1988; Kelso et al., 1984; Shaiman, 1989) reflecting the presence of afferent dependent mechanisms in the control of speech movements. The distributed compensatory response to external perturbations is a direct reflection of the overall functional organization of the speech motor control process and is comparable to other sensorimotor actions observed for other motor behaviors such as postural adjustments (Marsden, Merton, & Morton, 1981; Nashner & Cordo, 1981; Nashner, Woollacott, & Tuma, 1979), eye-head interactions (Bizzi, Kalil, & Taglaisco, 1971; Morasso, Bizzi, & Dichgans, 1973), wrist-thumb actions (Traub, Rothwell, & Marsden, 1980), and thumb-finger coordination (Cole, Gracco, & Abbs, 1984). Changing the size of the oral cavity with the placement of a block between the teeth similarly results in compensatory changes in articulatory actions resulting in perceptually-acceptable vowel sounds (Lindblom, Lubker, & Gay, 1979; Fowler & Turvey, 1980). It appears that the speech motor control system is designed to achieve functional behaviors through interaction of ascending sensory signals with descending motor commands.

Human and nonhuman studies have shown that sensory receptors located throughout the vocal tract are sufficient to provide a range of dynamic and static information which can be used to signal position, speed, and location of physiological structures on a movement to movement basis (cf. Munger & Halata, 1983; Dubner, Sessle, & Storey, 1978; Kubota, Nakamura, & Schumacher, 1980; Landgren & Olsson, 1982 for reviews). Studies utilizing perturbation of speech motor output indicate that the rich supply of orofacial somatic sensory afferents have the requisite properties to interact with central motor operations to yield the flexible speech motor patterns associated with oral communication (Abbs & Gracco, 1984; Gracco & Abbs, 1985; Gracco & Abbs, 1988; Kelso et al., 1984). Because of the constantly changing peripheral conditions during speaking, the absolute position of vocal tract structures can vary widely depending on the surrounding phonetic environment. The speech motor control system apparently adjusts for these movement to movement variations by incorporating somatic sensory information from the various muscle and mechanoreceptors located throughout the vocal tract. Considerations outlined elsewhere (Gracco, 1987; Gracco & Abbs, 1987) suggest that the speech motor control system appears to use somatic sensory information in two distinct ways; in a comparative manner to feed back information on the attainment of a speech goal and to predictively parameterize or adjust upcoming control actions. Structurally, there is strong evidence for the interaction of sensory information from receptors located within the vocal tract with speech motor output at many if not all levels of the neuraxis (cf. Gracco, 1987; Gracco & Abbs, 1987 for a summary of the vocal tract representation in multiple cortical and subcortical sensory and motor regions). Further, brain stem organization, evidenced by reflex studies, demonstrate a range of complex interactions in which sensory input from one structure such as the jaw or face is potentially able to modify motor output from lip and tongue as well as jaw muscles (Bratzlavsky, 1976; Dubner et al., 1978; Smith, Moore, Weber, McFarland, & Moon, 1985; Weber & Smith, 1987). It appears that there are multiple synaptic interactions possible throughout the neural system controlling the vocal tract, with the specific interaction dependent on how the system is actively configured.

Speech motor actions involve the activation or inactivation of various muscles of the vocal tract which are adjusted based on the peripheral conditions and the specific phonetic requirements. An important question related to the neural representation for speech is the character of the underlying activation process for different articulatory actions. A number of recent studies, evaluating the kinematic characteristics of different articulators, are consistent with a single sensorimotor process to generate a variety of articulatory actions. One method for evaluating the similarity in the underlying representation for multiple speech sounds and their associated movement dynamics is to compare the geometric (normalized) form of velocity profiles. A change in velocity profile shape accompanying experimental manipulation of phonetic context suggests a change in the movement dynamics, and by inference a change in the underlying neural representation. Conversely, a demonstration of trajectory invariance or scalar equivalence for a variety of movements suggests that different movements can be produced from the same underlying dynamics (Atkeson & Hollerbach, 1985; Hollerbach & Flash, 1982). That is, in order to produce movement variations appropriate to peripheral conditions and task requirements, it may be necessary only to scale the parameters of a single underlying dynamical relation: a much simpler task and, by inference, a simpler neural process. For movements of the vocal folds, tongue, lips, and jaw during speech it has been shown that changes in movement duration and to a lesser extent movement amplitude reflect a scaling of a base velocity profile (Gracco, submitted; Munhall, Ostry, & Parush, 1985; Ostry & Cooke, 1987; Ostry, Cooke, & Munhall, 1987; Ostry & Munhall, 1985). A scalar relation across a class of speech sounds involving the same articulators maintained for different initial conditions (different vowel contexts) suggests that the neural representation has been maximized and such a representation might reflect a basic component of speech production. That is, all speech movements may involve a simple scaling of a single characteristic dynamic (force-time) relationship (Kelso & Tuller, 1984) with the kinematic variations reflecting the influence of biomechanical and timing specifications. In addition, specification of control signals in terms of dynamics eliminates the need to specify individual movement trajectories since the path taken by any articulator is a consequence of the dynamics rather than being explicitly specified (see Kelso et al., 1984; Saltzman, 1986; Saltzman & Munhall, 1989). The scaling of individual actions appears to be another characteristic process that eliminates the need to store all possible phonetic variations explicitly. Rather, the

control process is a scaling of characteristic motor patterns adjusted for endogenous conditions (speaking rate, emphasis, upcoming functional requirements) and the surrounding phonetic environment (sensorimotor adjustments). The classic central-peripheral, motor program-reflex perspectives have given way to more reasonable and realistic issues including when and how sensory information may be used and how the different representations are coded for the generation of all possible speech movements.

Movement sequencing

A significant characteristic of many motor behaviors such as speech, locomotion, chewing, and typing is the production of sequential movements. Observations that interarticulator timing is not disrupted following perturbation (Gracco & Abbs, 1988), that speech rate can be modulated by changes in sensory input (Gracco & Abbs, 1989), and that perturbation induces minimal changes in speech movement duration (Gracco & Abbs, 1988; Lindblom et al., 1987) are consistent with an underlying oscillatory mechanism for speech. Further, somatic sensory-induced changes in the timing of oral closing action (due to lower lip perturbation) is consistent with an underlying oscillatory process (Gracco & Abbs, 1988; 1989). Qualitative observations of temporal consistency of sequential movements are also consistent with an underlying oscillatory or rhythm generating mechanism. Presented in Figure 4 are 24 superimposed movements of the upper lip, lower lip, and jaw for the sentence "Buy Bobby a Poppy."

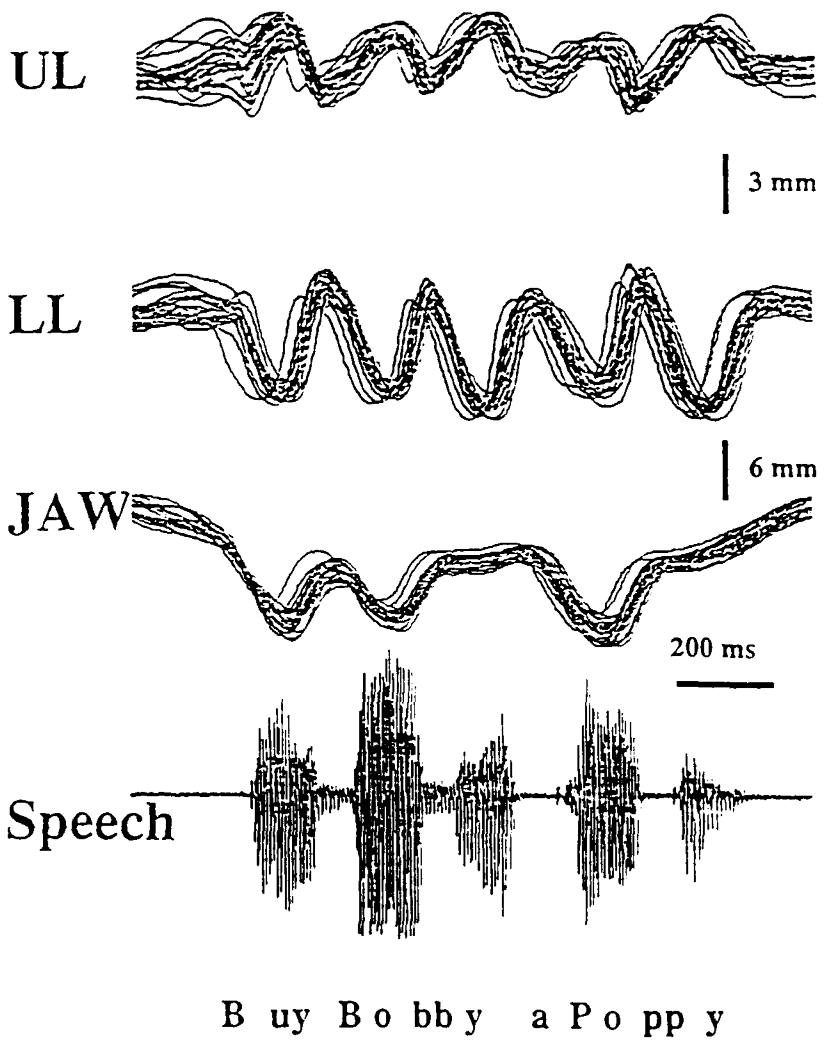

Figure 4. **Superimposed upper lip (UL), lower lip (LL) and jaw (J) movements associated with 24 repetitions of the sentence "Buy Bobby a Poppy"; the patterns are remarkably similar displaying little spatiotemporal variation. Only the acoustic signal from a single repetition is shown.**

These repetitions were produced as part of a larger study and were produced at different times during the experiment. The subject produced one repetition per breath and each repetition was produced at a comfortable subject-defined rate. As can be seen there is a consistency to the repetitions that suggests an underlying periodicity indicative of a rhythmic process. A few studies, attempting to address the periodicity and apparent rhythmicity of speech have demonstrated the presence of some form of underlying frequency generating mechanism. Ohala (1975) recorded over 10,000 jaw movements within a 1.5 hour period of oral reading and was able to identify frequencies ranging from 2-6 Hz with significant durational variability. Kelso et al. (1985) using reiterant productions of the syllable "ba" or "ma" demonstrated a rather strong periodicity at approximately 5-6 Hz with minimal durational variability. The findings of the Kelso et al., (1985) are consistent with an underlying oscillatory process. In contrast, the range of frequencies found by Ohala (1975) may reflect the frequency modulation associated with the sounds of the language, a factor minimized in the Kelso et al. (1985) study. The modulation of frequency, dependent on specific aerodynamic properties of the specific sounds and surrounding articulatory environment may be a mechanism underlying the speech movement sequencing (see also Saltzman & Munhall, 1989 for further discussion of serial dynamics). That fact that the frequency values reported by Kelso et al. (1985) were similar for "ba" and "ma" suggest that vowels may be a major factor in determining the local periodicity. However, it is the case that the individual movements or movement cycles are not the same; local frequencies are different depending on the phonetic context.

In addition, speech production involves many of the same muscles as such automatic behaviors as breathing, chewing, sucking, and swallowing. It has been suggested that the mechanisms underlying speech may incorporate, to some degree, the same mechanisms as more automatic motor behaviors but adapted for the specialized function of communication (Evarts, 1982 Gracco & Abbs, 1988; Grillner, 1982; Kelso, Tuller, & Harris, 1983; Lund, Appenteng, & Seguin 1982). Few studies have focused specifically on the similarity of speech with more innate, rhythmic motor behaviors (Moore, Smith, & Ringel, 1988; Ostry & Flanagan, 1989) with mixed interpretations. Recent experiments and theoretical perspectives on the organization of central pattern generators for rhythmic behaviors such as locomotion, respiration and mastication suggest a more flexible conceptualization of the possible behavioral outputs than has previously been envisioned for the neural control of rhythmic behaviors (see Cohen, Rossignol, & Grillner, 1988; Getting, 1989 for reviews). For example, *in vitro* results suggest that the central pattern generator for respiration may more appropriately be considered as two separate but interrelated functions; one generating the rhythm and one generating the motor pattern (Feldman, Smith, McCrimmon, Ellenberger, & Speck, 1988). The implication for other rhythmic and quasi-rhythmic behaviors such as speech, is that each function can be modulated independently thus generalizing the concept of a central pattern generator to a wider range of behaviors. Recently, Patla (1988) has suggested that nonlinear conservative oscillators are the most plausible class of biological oscillators to model central pattern generators in that they provide the necessary time-keeping function as well as independent shaping of the output (see also Kelso & Tuller, 1984). The recent demonstration by Moore et al. (1988) that mandibular muscle actions for speech are fundamentally different than for chewing suggests that the patterning for each behavior is different. That is, speech and chewing may share the same generator but have different patterning or, conversely, rely on different generators and patterns. Conceptually and theoretically, a fundamental frequency oscillator and static nonlinear shaping function can generate a number of complex patterns. While speculative, some current CPG models have the necessary complexity to be tentatively applied and rigorously tested as to their appropriateness for speech motor control.

SUMMARY

From the present perspective, the speech motor control system is viewed as a biophysical structure with unique configurational characteristics. The structure does not constrain the systems' operation but significantly affects the observable behavior and hence the resulting acoustic manifestations. Consideration of the structural organization and the potential contributions from biomechanical interactions are suggested as potential explanations for some speech motor variability. Sensorimotor mechanisms were implicated as the means by which adjustments in characteristic vocal tract shapes can be dynamically and predictively modified to

accommodate the changing peripheral conditions. From the perspective of the vocal tract as the controlled system, the consistent coordinative timing relationships reflect the functional modification of all the control elements or articulatory structures. Rather than describing sound production as the modulation or assembly of discrete units of action, the current functional perspective suggests that entire vocal tract actions are modulated to regulate acoustic/aerodynamic output parameters. The different parameters are realized by manipulation of the frequency of the forcing function applied uniformly to the control elements of the system. Rather than a parametric forcing in which some parameter such as stiffness is viewed as a regulated variable, it is hypothesized that the system is extrinsically forced by manipulation of the frequency of neural output consistent with the spatial requirements (e.g. movement extent) of the task. The frequency-modulated neuromotor actions are then filtered through a complex peripheral biomechanical environment resulting in elaborate kinematic patterns. Speech motor control is viewed as a hierarchically organized control structure in which peripheral somatic sensory information interacts with central motor representations. The control scheme is viewed as hierarchical from the standpoint that the motor adjustments are embedded within a number of levels of organization reflecting the overall goal of the motor act, communication. Modifications in the control signals reflect the parallel processing of multiple brain regions to scale and sequence changes in overall vocal tract states (Gracco & Abbs, 1987). The organizational characteristics of speech as a motor control system are fundamentally similar to other sequential motor actions and are felt to involve a limited number of general sensorimotor control processes.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscles responses to load perturbation of the lip during speech. *Journal of Neurophysiology, 51*, 705-723.

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multi-movement coordination: Sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior, 16*, 195-232.

Atkeson, C. G., & Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience, 5*, 2318-2330.

Bell-Berti, F., & Hirose, H. (1973). Stop consonant voicing and pharyngeal cavity size. *Journal of the Acoustical Society of America, 53*, 295-315.

Bernstein, N. (1967). *The co-ordination and regulation of movements.* New York: Pergamon Press.

Bizzi, E., Kalil, R., & Tagliasco, V. (1971). Eye-head coordination in monkeys: Evidence for centrally patterned organization. *Science, 173*, 452-454.

Boschert, J., Hink, R. F., & Deecke, L. (1983). Finger versus toe movement-related potentials: Further evidence for supplementary motor area (SMA) participation prior to voluntary action. *Experimental Brain Research, 52*, 73-80.

Bratzlavsky, M. (1976). Human brainstem reflexes. In M. Shahani (Ed.), *The motor system: Neurophysiology and muscle mechanisms.* New York: Elsevier.

Brown, W. S., & McGlone, R. E. (1979). Supraglottic air pressure during a valsalva maneuver. In J. J. Wolf & D. H. Klatt, (Eds.), *Speech Communication Papers: 97th Meeting of the Acoustical Society of America*, New York.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22*, 129-159.

Cohen, A. H., Rossignol, S., & Grillner, S. (1988). *Neural control of rhythmic movements in vertebrates.* New York: Wiley.

Cole, K. J., Gracco, V. L., & Abbs, J. H. (1984). Autogenic and nonautogenic sensorimotor actions in the control of multiarticulate hand movements. *Experimental Brain Research, 56*, 582-585.

Curry, S. H., Peters, J. F., & Weinberg, H. (1978). Choice of active electrode site and recording montage as variables affecting CNV amplitude preceding speech. In D. A. Otto (Ed)., *Multidisciplinary perspectives in event-related brain potential research* (pp. 275-279). Washington, DC: US Environmental Protection Agency.

Deecke, L., Scheid, P., & Kornhuber, H. H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research, 7*, 158-168.

Dubner, R., Sessle, B. J., & Storey, A. T. (1978). *The neural basis of oral and facial function.* New York: Plenum Press.

Evarts, E. V. (1982). Analogies between central motor programs for speech and for limb movements. In S. Grillner, B. Lindblom, J. Lubker, & A Persson (Eds.), *Speech motor control* (pp. 19-41). Oxford: Pergammon Press.

Evarts, E. V., Bizzi, E., Burke, R. E., DeLong, M., & Thach, W. T., (Eds.) (1972). *Neurosciences Research Symposium Summaries, Vol. 6: Central control of movement* (pp. 1-170).

Feldman, J. L., Smith, J. C., McCrimmon, D. R., Ellenberger, H. H., & Speck, D. F. (1988). Generation of respiratory patterns in mammals. In A. H. Cohen, S. Rossignol, & S. Grillner (Eds.), *Neural control of rhythmic movements in vertebrates* (pp. 73-100). New York: Wiley and Sons.

Folkins, J. W. (1976). *Multidimensional lower lip displacement resulting from activation of individual labial muscles: Development of a static model.* Doctoral dissertation, University of Washington.

Folkins, J. W., & Zimmermann, G. N. (1981). Jaw-muscle activity during speech with the mandible fixed. *Journal of the Acoustical Society of America, 69*, 1441-1445.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research, 18*, 207-220.

Fowler, C. A. (1977). *Timing control in speech production.* Indiana University Linguistics Club, Bloomington, IN.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (pp. 373-420). New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Journal of Phonetics, 37*, 306-326.

Fowler, C. A., Gracco, V. L., & V.-Bateson, E., (1989). Remote and local effects of stress within and among articulatory subsystems. *Journal of the Acoustical Society of America, 86(1)*, S115.

Fromkin, V. A. (1966). Neuro-muscular specification of linguistic units. *Language & Speech, 9*, 170-199.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. (1971). Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organization of certain biological systems* (pp. 329-345). Cambridge, Mass: MIT Press.

Getting, P. A. (1989). Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience*, 12, 185-204.

Gracco, V. L. (submitted). Some organizational principles for speech movement control. *Journal of Speech and Hearing Research*.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8*, 4628-4634.

Gracco, V. L. (1987). A multilevel control model for speech motor activity. In H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 57-76). Wien: Springer-Verlag.

Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysics, 54*, 418-432.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research, 71*, 515-526.

Gracco, V. L., & Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research, 75*, 586-598.

Gracco, V. L., & Abbs, J. H. (1987). Programming and execution processes of speech movement control: Potential neural correlates. In E. Keller & M. Gopnik (Eds.), *Symposium on motor and sensory language processes* (pp. 163-201). Hillsdale, NJ: Lawrence Erlbaum Associates,

Gracco, V. L., & Löfqvist (1989). Speech movement coordination: Oral-laryngeal interactions. *Journal of the Acoustical Society of American, 86*, S114.

Greene, P. H. (1972). Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (pp. 303-338). New York: Academic Press.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A Persson (Eds.), *Speech motor control* (pp. 217-230). Oxford: Pergammon Press.

Harris, K. S., Lysaught, G. F., & Schvey, M. M. (1965). Some aspects of the production of oral and nasal labial stops. *Language and Speech, 8*, 135-147.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Ho, T. P., Azar, K., Weinstein, S. & Bowley, W. W. (1982). Physical properties of human lips: Experimental and theoretical analysis. *Journal of Biomechanics, 15*, 859-866.

Hollerbach, J. M. & Flash, T. (1982). Dynamic interactions between limb segments during planar arm movement. *Biological Cybernetics, 44*, 67-77.

Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. *American Journal of Physiology, 15*, R928-R935.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech* (pp. 137-173). New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B, V.-Bateson, E., & Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812-832.

Kelso, J. A. S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. In, H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 105-128). Berlin: Springer-Verlag.

Kots, Ya. M. (1975). *The organization of voluntary movement*. New York: Plenum Press.

Kubota, K., Nakamura, Y, & Schumacher, G. H. (1980). *Jaw position and jaw movement*. Berlin: Verlag Volk und Gesundeit.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach (Ed.), *Tutorials in motor behavior* (pp. 3-47). Amsterdam: Elsevier/North Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 5-78). New York: Wiley.

Landgren, S., & Olsson, K. A. (1982). Oral mechanoreceptors. In S. Grillner, B. Lindblom, J. Lubker, & A Persson (Eds.), *Speech motor control* (pp. 129-140). Oxford: Pergammon Press.

Larsen, B., Skinhøj, E., & Lassen, N. A. (1978). Variations in regional cortical blood flow in the right and left hemispheres during automatic speech. *Brain, 101*, 193-209.

Lindblom, B., Lubker, J, & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7*, 147-161.

Lindblom, B., Lubker, J., Gay, T., Lyberg, P., Branderal, P., & Holgren, K. (1987). The concept of target and speech timing. In R. Channon & L. Shockery (Eds.), *In honor of Ilse Lehiste* (pp. 161-181). Dordrecht, The Netherlands: Foris Publications.

Löfqvist, A., & Yoshioka, H. (1981). Interarticulator programming in obstruent production. *Phonetica, 38*, 21-34.

Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication, 3*, 279-289.

Lubker, J. F., & Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America, 47*, 625-633.

Lund, J. P., Appenteng, K., & Sequin, J. J. (1982). Analogies and common features in the speech and masticating systems. In S. Grillner, B. Lindblom, J. Lubker, & A Persson (Eds.), *Speech motor control* (pp. 231-245). Oxford: Pergammon Press.

Lynn, A. M. J., & Yemm, R. (1971). External forces required to move the mandible of relaxed human subjects. *Archives of Oral Biology, 16*, 1443-1447.

Marsden, C., Merton, P., & Morton, H. (1981). Human postural responses. *Brain, 104*, 513-534.

Minifie, F., Abbs, J. H., Tarlow, A., & Kwaterski, M. (1974). EMG activity within the pharynx during speech production. *Journal of Speech Hearing Research, 17*, 497-504.

Morasso, P., Bizzi, E., & Dichgans, J. (1973). Adjustments of saccade characteristics during head movements. *Experimental Brain Research, 16*, 492-500.

Moore, C. A., Smith, A., & Ringel, R. L. (1988). Task-specific organization of activity in human jaw muscles. *Journal of Speech Hearing Research, 31*, 670-680.

Munger, B. L., & Halata, Z. (1983). The sensory innervation of primate facial skin. I. Hairy skin. *Brain Research Reviews, 5*, 45-80.

Müller, E. M., & Brown, W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 4,pp. 317-389). Academic Press.

Munhall, K. G., D. J. Ostry, & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 457-474.

Nashner, L. M., & Cordo, P. J. (1981). Relation of automatic postural responses and reaction-time voluntary movements of human leg muscles. *Experimental Brain Research, 43*, 395-405.

Nashner, L. M., Woollacott, M., & Tuma, G. (1979). Organization of rapid responses to postural and locomotor-like perturbation of standing man. *Experimental Brain Research*, 36, 463-476.

Ohala, J. J (1975). The temporal regulation of speech. In G. Fant & M. A. A. Tatham, (Eds.), *Auditory analysis and perception of speech* (pp. 431-454). London: Academic, .

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage, (Ed.), *The production of speech* (pp. 189-216). New York: Springer-Verlag.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Ostry, D. J., & Flanagan, J. R. (1989). Human jaw movement in mastication and speech. *Archives of Oral Biology, 34*, 685-693.

Ostry, D. J., & J. D. Cooke (1987). Kinematic patterns in speech and limb movements. In, E. Keller & M. Gopnik, (Eds.), *Motor and sensory processes of language* (pp. 223-235). Hillsdale, NJ: Lawrence Erlbaum.

Ostry, D. J., Cooke, J. D., & Munhall, K. G. (1987). Velocity curves of human arm and speech movements. *Experimental Brain Research, 74*, 1-10.

Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77*, 640-648.

Patla, A. E. (1988). Analytic approaches to the study of outputs from central pattern generators. In A. H. Cohen, S., Rossignol, & Grillner, S. (Eds.), *Neural control of rhythmic movements in vertebrates* (pp. 455-486). New York: Wiley.

Perkell, J. S. (1969). Physiology of speech production: Results and implications of a quantitative cineradiographic study (Research Monograph No. 53). Cambridge, MA: MIT Press.

Perkel, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Progress Bulletin, 6(3)*, 223-344.

Perlman, A., Luschei, E., & DuMond, C. E. (1989). Electrical activity from the superior pharyngeal constrictor during reflexive and nonreflexive tasks. *Journal of Speech and Hearing Research, 32*, 749-754.

Saltzman, E. L. (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology, 20*, 91-163.

Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 129-144). Berlin: Springer-Verlag.

Saltzman, E. L. & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86*, 78-87.

Smith, A., Moore, C. A., Weber, C. M., McFarland, D. H., & Moon, J. B. (1985). Reflex responses of the human jaw-closing system depend on the locus of intraoral mechanical stimulation. *Experimental Neurology, 90*, 489-509.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America, 82*, 847-863.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research, 16*, 397-420.

Thelen, E. (1985). Developmental origins of motor coordination: Leg movement in human infants. *Developmental Psychology, 18*, 1-22.

Thelen, E. (1986). Treadmill-elicited stepping in seven-month-old infants. *Child Development*, 57, 1498-1506.

Traub, M. M., Rothwell, J. C., & Marsden, C. D. (1980). A grab reflex in the human hand. *Brain, 103*, 869-884.

Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America, 76*, 1030-1036.

Tuller, B., Kelso, J. A. S., & Harris , K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology, 8*, 460- 472.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Ed.), *Perceiving, acting and knowing: Toward an ecological psychology* (pp. 211-265). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Weber, C. M., & Smith, A. (1987). Reflex responses in human jaw, lip, and tongue muscles elicited by mechanical stimulation. *Journal of Speech and Hearing Research, 30*, 70-79.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*, 1322-1336.

FOOTNOTE

*To appear in G. Hammond (Ed.), *Cerebral control of speech and limb movements*. Amsterdam: Elsevier.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 27-44

Sensorimotor Mechanisms in Speech Motor Control*

Vincent L. Gracco

A conceptual model of speech motor control is developed in which the elemental units for speech are sound-producing coordinated movements of the vocal tract. The perspective taken is that the degrees of control freedom are at a system level; in the operation of the processes that implement the speech motor action. Speech motor control is conceptualized as a multistage parallel process in which vocal tract specifications are activated by central motor commands which interact with a central rhythmic output to produce serial coordinated movements required for sound generation. Vocal tract specifications include the selection of characteristic neuromotor patterns, which map isomorphically onto the phonemes of the language. Coordination of the contributing movements and on-line spatial adjustments within and among vocal tract structures are inherent in the neuromotor patterning and activation processes, respectively. The elemental units are retrievable patterns stored in the central nervous system and instantiated by the directed action of the posterior parietal cortex. Two major brain systems (basal ganglia-supplementary motor area and the cerebellar-premotor area), are proposed to play major roles in implementing neuromotor specifications by modulating the characteristic patterns and the sequencing their actions into larger meaningful units of production. It is the action and interaction of these sensorimotor mechanisms that result in the speech motor patterns characteristic of human verbal communication.

INTRODUCTION

> If you root yourself in the ground, you can afford to be stupid. But if you move, you must have mechanisms for moving, and mechanisms to ensure that the movement is not utterly arbitrary and independent of what is going on outside.
>
> —Patricia Smith Churchland (1986).

After years of theoretical debate and endless empirical investigations, the classic central-peripheral issue that has guided much of the research in motor theory has given way to the more reasonable perspective that movement reflects an interaction of peripheral influences and central motor processes; behavior is sensorimotor in nature. Moreover, it is becoming increasingly clear that any behavior is a reflection of multiple overlapping and interacting influences, each of which needs to be identified. The purpose of identifying the subcomponents is not strictly to assign function to structure but to evaluate their potential contribution to the overall process, and hence allow development of realistic and biologically plausible working models of the system. An important research focus in human motor behavior has become the development of models that capture the essence of sensorimotor control (P. M. Churchland, 1989; Marr, 1982; McClelland & Rumelhart, 1986; Pellionisz & Llinas, 1979; Pellionisz & Llinas, 1985; Rumelhart & McClelland, 1986). The rationale for such an endeavor is two fold: first, there is an inherent richness and intricacity to even the simplest problem of sensorimotor control, and second, an implicit assumption that higher functions such as cognition are not discontinuous with the lower level sensorimotor functions that implement them (see P. S. Churchland, 1986). In this regard a statement by Hughlings Jackson made over 115 years ago seems prophetic:

> I cannot conceive what even the highest nervous centres can possibly be, except developments out of lower nervous centres, which no one doubts to represent impressions and movements.
>
> —J. Hughlings Jackson (1875).

The writing of this paper was supported by NIH grants DC-00121 and DC-00594. Carol Fowler is acknowledged for helpful comments.

Because of its well-learned and ecologically significant nature, speech is an ideal behavior for the investigation of sensorimotor control mechanisms. Moreover, as a reflection of one of man's most highly developed behaviors, a thorough understanding of the processes of communication may provide valuable insight into the operation and functional organization of the human nervous system.

The purpose of the following chapter is to propose a preliminary conceptual model of speech production from a functional (e.g., communicative) perspective that is grounded as much as possible in physiological mechanisms and plausible nervous system processes. Implicit in any model of human behavior is the tacit assumption that the hypothetical processes or functions actually exist in some form in the central nervous system or at least emerge from central, peripheral and/or biomechanical interactions. As such, the conceptual model will be limited to constructs known or suspected from nervous system mechanisms. How many different mechanisms are required to explain the observable behavior? What aspects of the observable behavior need to be explained or accounted for? What role does peripheral sensory information play in the control of speech movements? What are the organizational principles for speech production? These are some of the issues that will be dealt with in the following chapter. Because the model presented is conceptual in nature and preliminary in form, only basic principles will be presented and many details will be lacking. One important component that will not be discussed is the contribution of the biomechanical periphery to the shaping of the complex kinematic patterns characteristic of speech. Only through incorporation of the physical properties of the vocal tract with underlying sensorimotor mechanisms can a realistic and parsimonious model be constructed. Within this limitation, a focus on underlying global sensorimotor processes should provide an additional and potentially viable perspective on speech production and perhaps a better perspective on motor speech disorders as well.

Organizational structure for speech motor control

In order to discuss the sensorimotor mechanisms that may underlie speech production it is first necessary to determine the most plausible conceptualization of the system being controlled. During speech, different vocal tract actions are sequenced to produce groups of linguistically-relevant sounds. Over the last 8-10 years, attempts have been made to determine the specific organization for speech motor control, i.e., to identify the appropriate level of articulatory organization. The lack of invariant individual articulatory actions and the relatively consistent ensemble articulatory actions suggests that the nervous system does not explicitly control the action of a single muscle or articulator (Gracco & Abbs, 1986; Kelso & Tuller, 1984; Saltzman, 1986). Rather, speech motor actions are organized at a level that reflects the interaction of a number of muscles and/or articulators engaged in the same functional task. For example, the final positions of the upper lip, lower lip, and jaw during bilabial production are not invariantly attained but vary systematically within some limit such that an apparent goal, oral closure, is achieved (Gracco & Abbs, 1986). Similarly, when the movement of an articulator is unexpectedly impeded during its normal motion, displacement is increased in the perturbed articulator as well as in various unperturbed articulators actively involved in producing the movement goal (Abbs & Gracco, 1984; Gracco & Abbs, 1985; 1988; Kelso, Tuller, V.-Bateson, & Fowler, 1984; Shaiman, 1989). Relative timing patterns observed for the upper lip, lower lip, jaw, and lower lip, jaw, and larynx in various phonetic contexts suggests that coordinative adjustments across vocal tract components is an important property of the motor control process (Gracco & Löfqvist, 1989; Gracco, 1988; Gracco & Abbs, 1986; Löfqvist & Yoshioka, 1981;1984). Consistent relative timing relations, distributed compensatory actions, and systematically variable articulatory interactions suggest that speech motor control must be viewed from a perspective encompassing ensemble articulatory actions. An important research question is the size of the ensemble, i.e., the size of the production unit.

One possible approach to the question of articulatory organization is captured in the construct of a coordinative structure (Fowler, Rubin, Remez, & Turvey, 1986; Kelso, 1986; Kugler et al., 1980; Saltzman & Kelso, 1987; Turvey, 1977). For speech, such a style of organization involves a number of flexible, but relatively constrained articulatory actions or ensembles, represented conceptually as tract variables (see Saltzman, 1986; Saltzman & Munhall, 1989) or physiologically as functional synergies (Fowler et al., 1980; Kelso, 1986; Kelso & Tuller, 1984) assembled into larger action units to produce sound (Browman & Goldstein, 1989; 1990; Saltzman, 1986; Saltzman & Munhall, 1989). From this perspective, speech sounds result from

the assembly of vocal tract actions (constriction producing events) from presumably independent primitive gestural units (Browman & Goldstein, 1989; Fowler et al., 1980; Kelso, 1986). This particular organizational scheme can be thought of as horizontal in the sense that the vocal tract is partitioned into articulatory subsystems which are marshalled into task-specific patterns (Kelso, 1986). However, one assumption of the coordinative structure, i.e., an active process that pieces together or assembles elementary or primitive articulatory actions has never been critically evaluated.

The construction of complex behaviors from simpler movements has been suggested for other tasks such as locomotion (Flashner, Beuter, & Arabyan, 1988), handwriting (Hollerbach, 1981: Morasso & Mussa-Ivaldi, 1982; Edelman & Flash, 1987; Lacquaniti, 1989) and pointing movements (Atkeson & Hollerbach, 1985: Morasso, 1981). A major difference, however, is that the behavior is organized vertically in the sense that complex behavioral sequences are composed of a smaller segments involving the entire effector unit rather than anatomical parts. For example, a primitive stroke in handwriting would involve all necessary components of the shoulder, arm and hand to produce a curved line (an elemental stroke) rather than isolated actions of the parts. Using the same analogy, speech production may be described as the concatenation of fundamental actions such as opening the vocal tract (as in the production of vowels) and closing the tract (as in the production of consonants) which produce or modulate sound. Rather than viewing the production of a /p/, for example, as involving a number of independent gestures (lip aperture gesture, a glottal gesture, an oral and pharyngeal gesture, and a velar gesture) assembled through a coordinative process, a simpler perspective is to view speech production in a wholistic sense in which characteristic neuromotor patterns, involving all components of the vocal tract, is the elemental control structure for speech. It can be argued that observations of distributed compensatory actions involving local and remote articulatory adjustments (Abbs & Gracco, 1984; Folkins & Abbs, 1975; Folkins & Zimmermann, 1982; Gracco & Abbs, 1988; Kelso, et al., 1984; Shaiman, 1989) are consistent with a level of organization in which vocal tract configurations are manipulated with no need for additional processes to assemble fundamental, nonspeech producing units. Similarly, recent findings such as the apparent adjustment in laryngeal timing to lower lip perturbation (Munhall, Löfqvist, & Kelso, in press) and the consistent relative timing among lip constriction/occlusion movements and glottal devoicing (see Figure 1 from Gracco & Löfqvist, 1989) suggest that neuromuscular adjustments across vocal tract structures are accomplished through manipulation of a common driving signal (Gracco, 1988) applied in a systematic manner to all active components of the vocal tract involved in producing a particular sound. It is apparent, however, that the available empirical evidence is consistent with either perspective and that conceptually identification of "the" primitive units of speech motor control is not important. Only in attempting to develop a realistic and parsimonious neurobiological and biophysical model of speech motor control does this issue has direct theoretical relevance.

CHARACTERISTIC MOTOR PATTERNS

As suggested above, coordinated sound-producing vocal tract actions, consistent with a segmental organization, are viewed as the smallest functioning structural units in the sensorimotor control process for speech. These hypothesized units are not abstractions, but characteristic neuromotor patterns whose implementation result in the production of sound. The characteristic patterns are similar to ideas presented by others such as Joos (1948), Fowler (1983), Saltzman and Munhall (1989) and Löfqvist (1990) but differ mainly in their level of description. At a neurophysiological level, these characteristic patterns are not invariant but are hypothesized to reflect a reference neural substrate which other sensorimotor processes act on resulting in output variability. This conceptualization is different from earlier speech production models which postulated the presence of invariant motor commands in that the patterns are one part of a distributed process, not the output of the system. The suggestion that speech production involves characteristic (not invariant) patterns is both logical and observable. For example, bilabial production always involves, to some degree, the same muscles produced with related characteristic actions. For example, presented in Figure 2a and 2b is a representative neuromuscular pattern for the upper and lower lip muscles and the resulting movement for the nonsense word "sapapple." Within certain boundary conditions, oral opening for an open vowel for /ae/ will result in some activity in upper lip and lower lip elevator and depressor muscles, respectively indicated in the figure (Figure 2a) by levator labii superior (LLS) and depressor labii inferior (DLI) (Figure 2b).

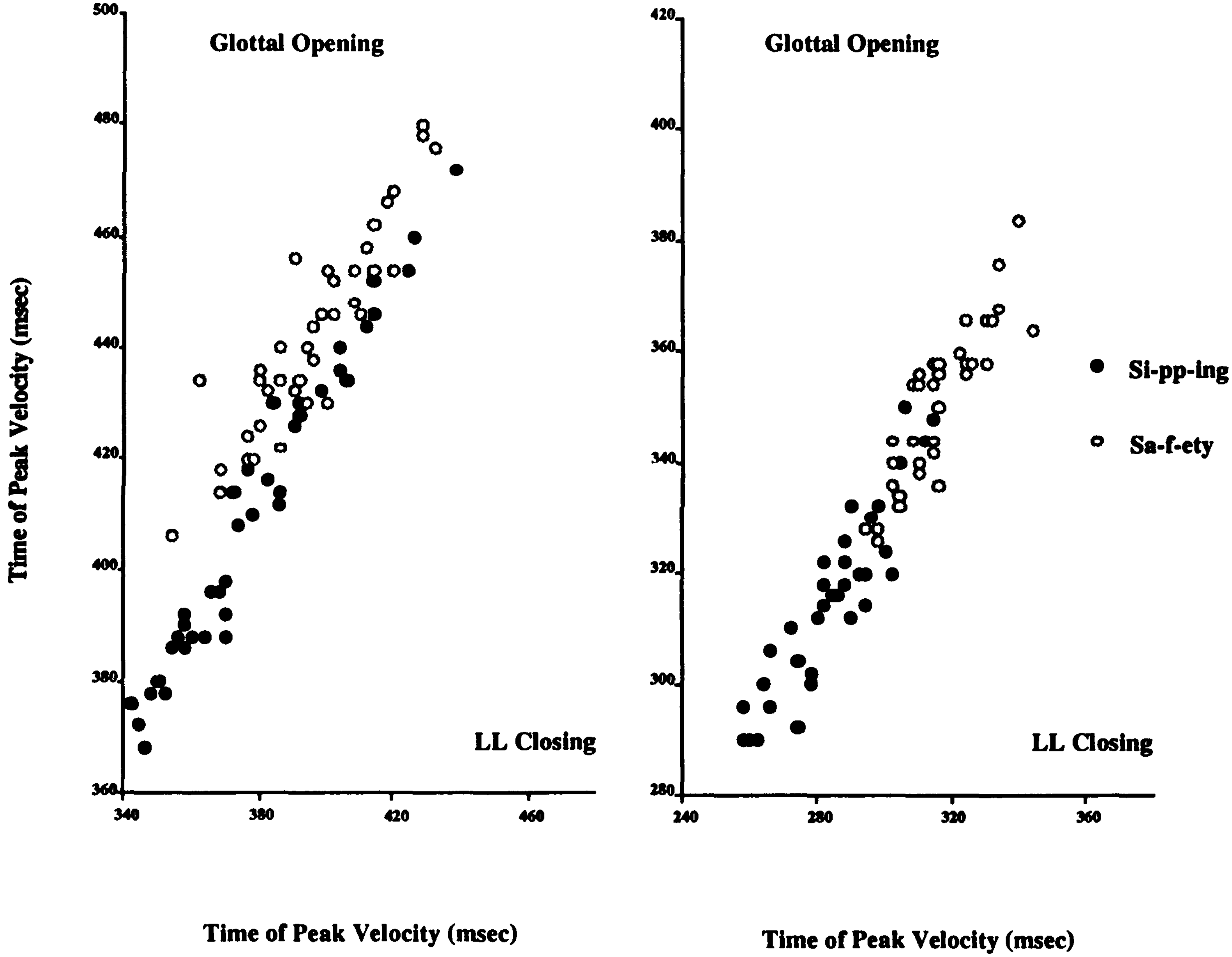

Figure 1. Scatterplots of the relative timing of lower lip peak velocity and the glottal devoicing peak velocity for occlusion /p/ (filled circles) and frication /f/ (open circles) for two subjects. For both subjects the relative timing of the articulatory events is constrained and similar across manner of production.

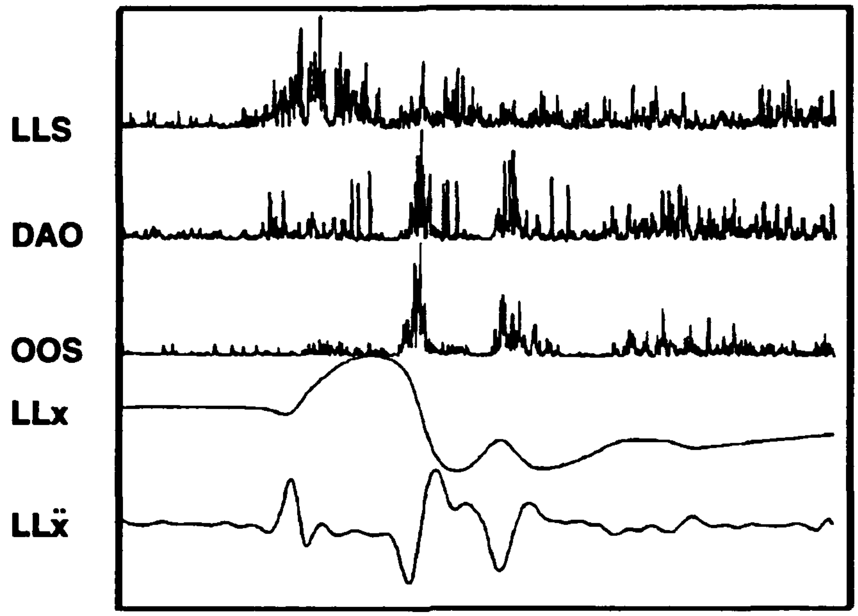

Figure 2a. **Rectified muscle activity for an upper lip elevator (levator labii superior, LLS), two upper lip depressors (depressor anguli oris, DAO, and orbicularis oris superior, OOS), upper lip displacement (ULx) and acceleration (bottom trace) illustrating a portion of what can be considered a neuromuscular pattern for oral closing. For the upper lip, the large negative-going acceleration marks the onset of the segment, followed by phasic bursts of muscle activity in DAO and OOS accompanying the oral closing.**

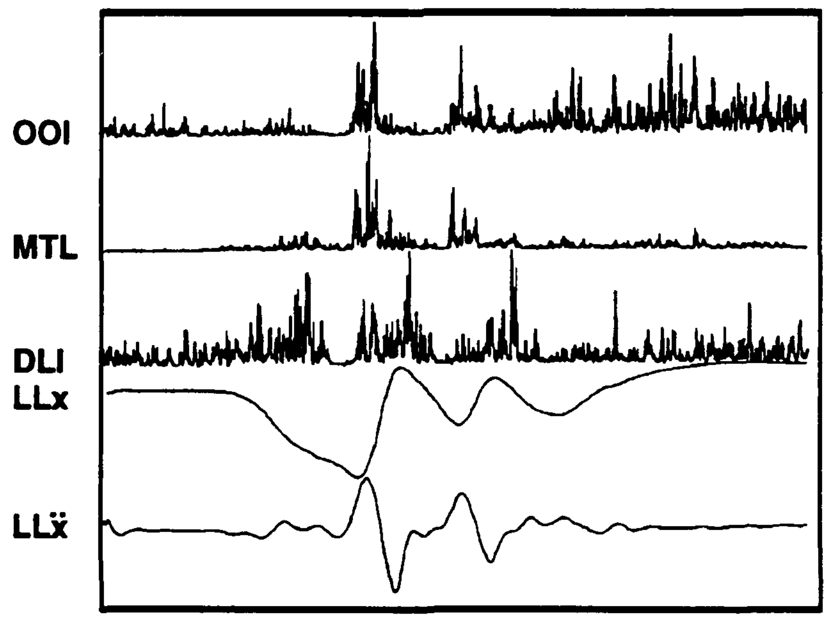

Figure 2b. **Rectified muscle activity for an lower lip depressor (depressor labii inferior, DLI), two lower lip elevators (orbicularis oris inferior, OOI, and, mentalis, MTL), lower lip displacement (LLx) and acceleration (bottom trace). For the lower lip, the large positive-going acceleration marks the onset of the segment, followed by phasic bursts of muscle activity in OOI and MTL accompanying the oral closing. This pattern of activation, along with the one presented in 2a, are considered characteristic of all bilabial sounds.**

Oral closing for any bilabial will involve some degree of activity in upper lip depressor muscles such as depressor anguli oris (DAO) and orbicularis oris superior (OOS)(Figure 2a), and activity in lower lip elevators (Figure 2b) such as mentalis (MTL) and orbicularis oris inferior (OOI); some cocontraction in LLS and DLI will accompany the closing action presumably to increase the overall stiffness of the lips and/or perhaps to damp the movements. This description reflects a consistent pattern of muscle action that accompanies all bilabial sounds. In contrast, bilabials are not produced with the tongue and, all things equal, are usually produced at a faster rate than vowel sounds. While there are certainly differences in some of the other contributing muscles in the vocal tract depending on whether the sound is /p/, /b/, or /m/, these are based on the particular aerodynamic or acoustic requirements for the sound. Similarly, the relative timing of such actions are also systematically related indicating that while the timing patterns may differ, they are related in a predictable manner observing simple scaling laws (Gracco, submitted). What uniquely defines each sound in the language is its particular neuromuscular configuration reflecting a distinct spatio-temporal pattern of activation and resulting motion. These patterns are not designed to explain all the details of observable speech movement actions, but are viewed as one fundamental component in the motor control process. Each component or group of components in the specification may have different activation patterns which reflect the form of the signal that impinges on lower motor neurons. In part the activation patterns reflect the contribution of the specific articulator to the sound as well as adjustments for the different biomechanical properties of the articulators. The activation patterns for the lip and the jaw muscles, for example, reflect their contribution to closing the oral end of the acoustic tube; the activation patterns are phasic, producing rapid closing movements; and the timing of medial pterygoid action occurs before the labial muscles due to the inertia of the jaw. In contrast, the activation patterns for the pharyngeal constrictors are more tonic and of longer duration reflecting their role in adjusting the tissue impedance of the vocal tract walls. In this regard, these patterns are viewed functionally as representing the essential dynamics of speech movement production and modulated by the differential filtering properties of the biomechanical periphery.

Prior to motor output at the periphery, these characteristic patterns are proposed to have a two or three dimensional spatial representation within, at least, the primary motor cortex and perhaps other nonprimary motor areas as well. Rather than attempt to present a speculate schematic spatial representation within the central nervous system, a schematic of a characteristic neuromuscular implementation realized at the periphery will be presented. Shown in Figure 3 is a representation of the output signals sent to various neuromuscular components of the vocal tract to produce a /p/. Given that many of the details are not currently known, the figure provides only the important neuromuscular components of the pattern. Further, muscle actions are functional grouped such that upper lip depressors (orbicularis oris superior and depressor anguli oris), for example, are only represented based on their articulatory consequences. At this level of observation, the characteristic motor patterns are isomorphic with the gestural constellations in the computationally sophisticated Linguistic Gestural Model (LGM) developed and implemented at Haskins Laboratories by Browman, Goldstein, and colleagues (Browman & Goldstein, 1985, 1986, 1989, 1990) and incorporates the aspects of earlier and more recent properties of the task dynamic model (TD) developed and refined by Saltzman and colleagues (Saltzman, 1986; Saltzman & Kelso, 1987: Saltzman & Munhall, 1989). The major difference (besides the fact that the TD and LGM are computational and this model has no such constraints!) is that much of the details that coordinate task-related vocal tract actions and differentiate sounds of the language are incorporated into stored nervous system elements which effectively reduce the on-line computational complexity. The rationale for such an approach is that speech as a well-learned (or over learned) motor behavior, incorporates much of its operation into automatic sensorimotor functions.

Within the current model, the characteristic vocal tract configurations and the phonemes of the language are isomorphic. This requires 43 different vocal tract specifications each with its characteristic neuromuscular specifications retained in nervous system memory; 43 is certainly not a number that would tax nervous system storage or processing capabilities. However, it is not clear that this is the fundamental unit of production or that phonemes are an important organizational unit; rather, sound producing vocal tract actions are the lowest level of sensorimotor control. As such, included in Figure 3 are the neuromuscular signals preceding oral closing (associated with a generic vowel) since in most cases opening and

closing actions must be tightly coupled. From a sensorimotor perspective, a VC or CVC (opening-closing) organization is more appealing as a unit of speech motor control. As suggested above, inherent in each pattern is the temporal coordination among the constituent components.

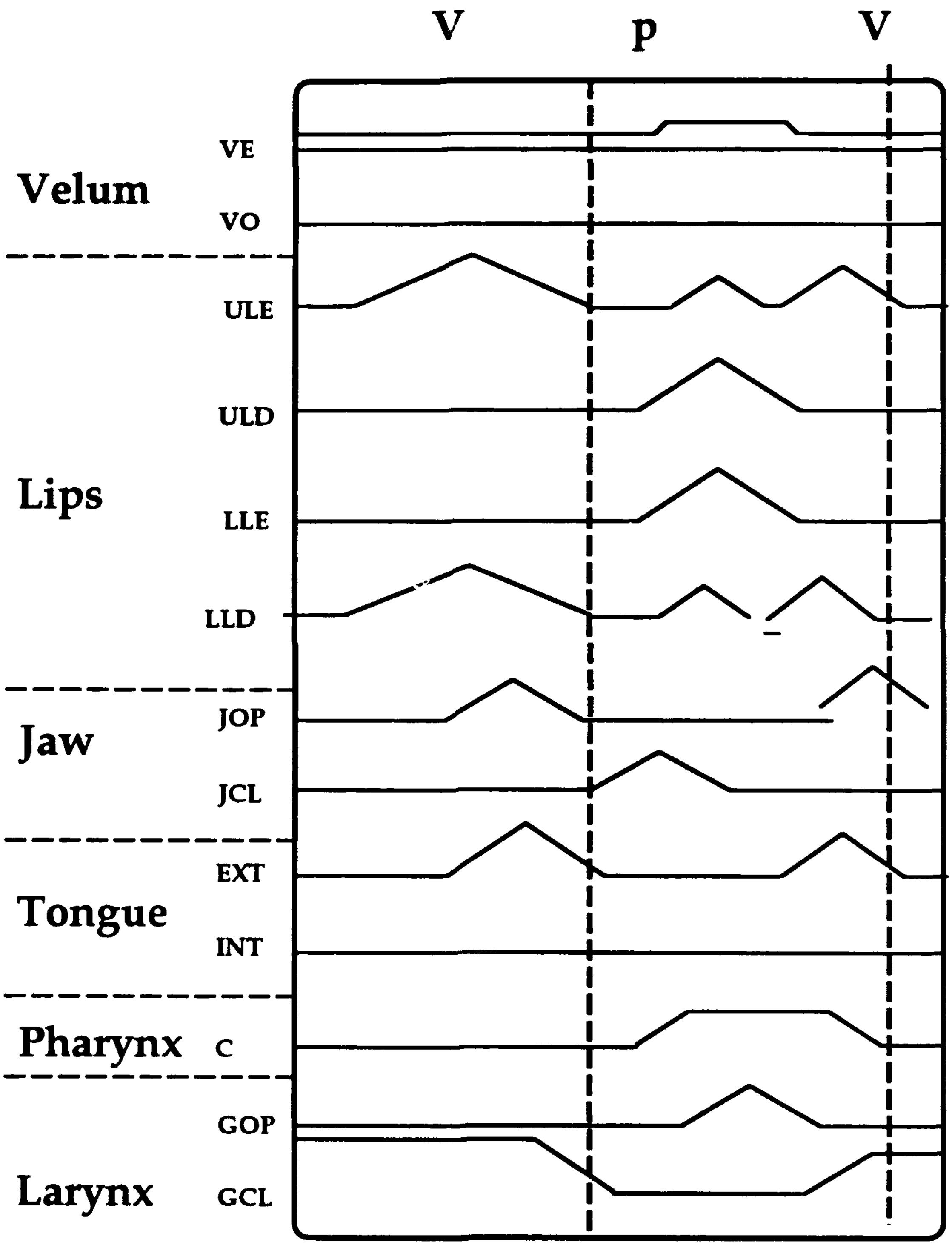

Figure 3. A schematic peripheral representation of a characteristic pattern of vocal tract activation for the bilabial /p/. The dotted lines generally demarcate the segment boundaries. Abbreviations are as follows: VE-velar elevator, VD-velar depressor, ULE-upper lip elevator, ULD-upper lip depressors, LLE-lower lip elevators, LLD-lower lip depressor, JOP-jaw openers, JCL-jaw closers, EXT-extrinsic (tongue muscles), INT-intrinsic, C-constrictors, GOP-glottal opener, GCL-glottal closers.

The time course of activation of the particular components and the particular signal shapes result in consistent and systematic coordinative patterns associated with various sounds. It is not surprising that relative timing is so consistent even in the face of mechanical perturbations (Gracco & Abbs, 1988; Gracco & Löfqvist, 1989; Gracco, 1988). These characteristic patterns can then be modulated according to other task related factors such as the distance to be moved, the overall rate of movement, and the presence of various stress adjustments. The patterns, with their inherent relative timing relations, can be easily compressed or expanded in a systematic manner by modulation of the frequency and/or amplitude of the input signals. It is also likely that the signal shapes vary for different articulators, since each articulator has specific biomechanical properties and such differences have generally been taken into account at least during development. Finally, separate processes extrinsic to the pattern such as those for speech rate and stress specifications should result in a unitary adjustment in all vocal tract structures. The observation of simultaneous respiratory, laryngeal, and oral adjustments accompanying emphatic stress-related manipulations is consistent with this organizational scheme (Fowler, Gracco, & V.-Bateson, 1989).

Before proceeding, a number of points should be discussed. First, while vocal tract specifications involve description of individual muscles and submuscle actions, it is not being suggested that the child learning to speak has to obtain control over all the individual muscular degrees of freedom. More likely, certain synergies exist, even at birth, that reflect constraints on the sound producing mechanism. As early as the birth cry, the infant is producing coordinated actions of the respiratory, laryngeal and supralaryngeal systems, or a cry would not be possible. As such, patterns are present that can be used as the basis for further differentiation. It is certainly plausible that these fundamental patterns are learned by the child during development based on some fundamental nonspeech actions emerging from breathing, sucking, chewing, swallowing, crying and early vocalizations. For example, breathing involves opening of the glottis during breathing which must be accompanied by relaxation (or significant reduction) in the activity of laryngeal adductors. Similarly, crying involves coordination of expiration with laryngeal adduction to produce vibration. As the child matures variations of this pattern may form the basis for voicing and devoicing. During chewing a basic pattern of jaw opening, accompanied by relaxation of jaw closing, forms a pattern that can be modified to produce the more variable jaw patterns for speech. Speech motor development may be envisioned as a learning process in which the child makes finer and more varied adjustments in its vocal tract, generalizing from fundamental nonspeech actions, to produced sounds. It is suggested that such actions become fixed once a sound is acquired by the child, and the characteristic neuromuscular pattern becomes a retrievable element in the child's sensorimotor repertoire.

There are a number of reasons for conceptualizing vocal tract actions from a neuromuscular perspective. First, the ability to fractionate control of muscles into functional chunks is consistent with the level of control exercised by the nervous system (English, 1982; Loeb, 1985). This is not to suggest that the nervous system controls muscles as opposed to movements; rather the detailed somatotopy and apparent fractionated control at the level of the motor cortex and brainstem can be exploited during speech acquisition to provide the framework to assemble patterns involving synergistic and part muscle actions. Second, description of the physiological characteristics of speech movements has the potential to provide a level of observation and detail not possible with more traditional kinematic accounts. This perspective captures the essence of the neural signals which co-occur with the contractile forces creating movement. With the concomitant development of realistic biomechanical models or elaboration of the biomechanical properties of the vocal tract, such signals can be used heuristically to determine which aspects of speech movement need to be explained in a control sense and which details emerge from passive biomechanical properties of the articulators. Finally, explicit consideration of the neuromuscular activation of vocal tract components provides insight into the manner in which these characteristic patterns become modified during implementation.

MODIFICATION OF VOCAL TRACT CONFIGURATIONS

To implement any action specific muscles involved can have only one of three distinct states of specification; activated, inhibited, or null. Unspecified articulators (null states) allow contiguous segmental vocal tract actions to intrude resulting in coarticulation (see Fowler, 1980; Kent & Minifie, 1977; Öhman, 1966; Saltzman & Munhall, 1989). Similarly, vocal tract actions involving the same articulator can be blended with

the rate of segmental adjustments determining the observable manifestation (see Munhall & Löfqvist, 1992; Stetson, 1951). Vocal tract actions may have contiguous phonetic segments with differing degrees of antagonistic action associated with a particular articulator. In certain contexts, neighboring submuscle actions of a particular articulator, such as the anterior and posterior portions of the tongue, may result in antagonistic action and articulator undershoot. One of the consequences of explicit consideration of neuromuscular organization is that coarticulation and other related phenomenon involving the smearing of characteristic vocal tract states should be affected by a combination of factors including degree of competition in contiguous segments and the overall speed or frequency of production. Further, if the sensorimotor control scheme outlined in the previous section is correct, there should be certain observations that are concomitant with coarticulatory phenomena. For example, if lip rounding is anticipated from a rounded vowel (/u/ for example) during the production of a nonlabial consonant such as /t/, the tongue body motion and resulting configuration for the /u/ should also show some affect of the intrusion of the /u/ segment. There should be an indication that the entire segment has blended rather than just a feature (see Daniloff & Hammarberg, 1973; Kent & Minifie, 1977 for reviews). In the present scheme, however, the specific coarticulatory influences can not be entirely predicted without a fundamental description and understanding of the neuromuscular configurations associated with specific vocal tract actions. This includes some understanding of the contribution of the biomechanical periphery and the interactions of the anatomical linkages to the sculpting of kinematic patterns (Gracco, 1990). In the following section, the role of peripheral sensory information will be considered as a means to modify the central motor commands.

Sensory influences

An important consideration concerning the sensorimotor control of speech is the influence of various sensory modalities. The specific extent and mode of sensory influences on speech motor output is still a matter of empirical investigation and theoretical contention and is one area that is often overlooked in speech production models. Information extracted from the different sensory modalities forms the basis for communicative, linguistic, or sensorimotor adjustments resulting in global as well as local effects on speech output. There are three sensory channels that have the potential to modify speech motor output each in overlapping but unique ways; visual, auditory, and somatic. During normal speaking situations, visual information regarding ones' vocal tract is not typically available; direct sensorimotor linkages are nonexistent. Rather, visual input is restricted to information regarding the communicative environment and provides what can be thought of as global influences on the motor control process. Faced with an environment that will require sound transmission across relatively long distances such as a classroom or lecture hall, the output intensity that a speaker uses will be adjusted to assure communincative effectiveness. Similarly, speaking to someone who is experiencing auditory acuity difficulties (temporary or permanent) the speaker may also modify the precision of articulatory adjustments to assist the listener. In general, visual information does not appear to play a significant or consistent role in the direct regulation of speech motor output. Rather, visual-motor influences can be thought of as adaptive and are more likely used for cognitive and certain linguistic adjustments affecting certain global sensorimotor parameters.

To evaluate the potential effects of auditory input on the motor control process, the auditory can be eliminated (temporarily) or distorted in various ways. Some useful information has been obtained using this kinds of experimental approach. For example, long duration exposure to high levels of auditory masking (Kelso & Tuller, 1983; Lane & Tranel, 1971; Ringel & Steer, 1963), delayed auditory feedback (Black, 1951; Fairbanks, 1955; Zimmermann, Brown, Kelso, Hurtig, & Forrest, 1988), and low pass filtering (Forrest, Abbas, & Zimmermann, 1986) are some of the conditions that can disrupt a subjects' auditory input. However, the issue of whether the modifications observed reflect the lack of auditory information or whether the modifications reflect long term exposure to novel feedback conditions has not been adequately addressed. Since sensory input can have both facilitatory and inhibitory effects on motor output introducing novel conditions for extended periods of time may result in changes that only indirectly, at best, reflect the potential contribution of the sensory modality to the normal motor control process. The best method for auditory disruption to date has been developed by Barlow and Abbs (1978) in which the subjects' own acoustic output (sidetone) is unpredictably eliminated for short durations (200 ms) on a small percentage of experimental trials. While such a paradigm does not provide a natural

probe into the system operation, it is much less obtrusive than previous techniques that suffer from potential adaptation effects.

Most researchers would agree that auditory information during speech development is critical to the acquisition of the sound patterns of the language. Long term elimination of auditory information or the lack of auditory information during speech development can severely affect the ability to maintain or acquire speech. As such, auditory input is considered instrumental in developing the characteristic neuromotor patterns that form the basis for the present model. Once acquired, however, the potential role of the auditory system may be limited. Even so, auditory information is still used in a corrective manner as evidenced by the adjustments one makes to slips of the tongue and other kinds of speech errors. In terms of on-line sensorimotor processes, reduced or distorted auditory information has been shown to result in rather subtle deficits in speech output. From some recent experimental evidence some have suggested that auditory information might play a role in the ongoing modulation of speech motor output (Barlow & Abbs, 1978; Forrest, Abbas, & Zimmermann, 1986; Zimmermann, Brown, Kelso, Hurtig, & Forrest, 1988). The dynamic properties of the acoustic signal can be related in a systematic, albeit nonlinear way, to articulatory motion, and could conceivably be useful in making predictive articulatory adjustments. To date, however, direct experimental evidence is limited.

Early research efforts to assess the potential role of somatic sensory information from skin and muscle receptors located throughout the vocal tract relied on local or nerve block anesthesia to eliminate sensory inflow. Results were equivocal but suggested to some that somatic sensory information, similar to auditory information, may play a role in speech acquisition but not in the regulation of the speech of adults (see Borden, 1979; Gracco & Abbs, 1987; Perkell, 1980 for reviews). It is doubtful, however, given the extent and degree of sensory innervation in the human vocal tract, that somatic sensory information can ever be truly eliminated. The lack of significant sensory reduction effects noted in some studies, then, suggests that speech can be produced, for a limited time without the full complement of incoming sensory information. This does not necessarily indicate that speech is afferent-independent, but that speech production is an integrated process with distributed and overlapping functions. Eliminating or reducing the contribution of one component of the process results in other components compensating for the loss.

More recently, mechanical loads unexpectedly applied to various articulators have been used to evaluate whether somatic sensory information is important to the ongoing motor control process. The reasoning is that, if sensory receptors located in various regions of the vocal tract are being continuously, or quasi-continuously, monitored during speaking, then disrupting articulatory movement should result in observable compensation. Results have clearly shown that somatic sensory signals have the necessary characteristics to be a useful in the on-line control of speech movements. Somatic sensory adjustments are rapid, usually less than a reaction time, and functionally organized such that the most directly perturbed articulators provide the major adjustment with secondary adjustments seen in anatomically remote functionally-related articulators. The distributed nature of the compensation strongly suggests that sensorimotor interactions, in the form of distributed synaptic linkages, are a feature of the neural organization for speech. Rapid, precise somatotopic and topographic adjustments have, to date, only been demonstrated from analysis of mechanical perturbation suggesting a dominant role for somatic sensory input in the ongoing modulation of speech motor output. This is not to suggest that other sensory modalities do not contribute to the ongoing sensorimotor control process; rather that the experimental evidence is lacking. It appears that the central nervous system is constantly receiving information on all phases of speech production and sensory considerations are as important in understanding motor control as perceptual considerations are important for understanding action.

Perhaps the best way to illustrate the manner in which direct sensory information can be used in the control of movement is to consider the motor task itself. Speaking involves the continuous modulation of the vocal tract producing local and global aerodynamic events structuring the air in characteristic ways. The specific vocal tract configurations are constantly changing during speaking with the same sound exhibiting variable movement patterns dependent on, among other things, phonetic context. From perturbation studies it is known that sensory information from somatic sensory receptors can interact with central motor commands to make short-term (within a few hundred milliseconds) and longer term contextual adjustments in speech motor output. The characteristic neuromuscular pattern previously

presented (Figure 3) can easily be adjusted through the vast sensorimotor linkages within and among vocal tract structures. As such, somatic sensory input from antecedent articulatory events can be used to modulate select properties of the neuromuscular pattern automatically (see Gracco, 1987 for discussion). In the case of a /p/ preceded by either a low vowel a neutral vowel or a high vowel, the oral aperture would reflect different degrees of openness with respect to some neutral or reference level. The somatic sensory input would, based on well established sensorimotor linkages, modulate the neuromotor pattern accordingly. Recent experimental results for bilabial sounds preceded by high or low vowels are consistent with the idea that there is an overall modulation of oral closing actions based on oral opening considerations (see also Folkins & Linville, 1983); an estimation of oral opening can be easily obtained from the jaw movement (or position) associated with the preceding vowel (cf. Gracco, 1987; Gracco, submitted). Further, when the oral opening distance is reduced due to a high vowel preceding closure, the upper and lower lip closing movements are reduced together suggesting that upper and lower lip control signals are modulated together. The resultant modulatory effects of sensorimotor linkages are dependent on a number of factors including the parameters of the central activation signals, and the strength and sign of the synaptic connections (the wiring). Sensorimotor interactions with characteristic neuromotor patterns provide a means to reduce the computational requirements of contextual variations by providing automatic adjustments in the control signals based on the conditions at the periphery.

SEQUENCING OF VOCAL TRACT ACTIONS

Speech is more than the specification of characteristic motor patterns adjusted for context. An important consideration in speech production is the sequencing of vocal tract actions into communicatively meaningful units of production. While speech is a specialized human function, the view taken here is that it is one of many important brain functions and any theoretical account must adhere to principles that are shared by other similar behaviors. If one accepts the premise that the human brain has evolved from earlier brains, (based on the need to predict and control species-specific events in the environment), then supposing that more complex, higher-level behaviors developed from lower level related behaviors, within and across species, is a logical extension. This is not to suggest that speech, locomotion, and handwriting, as examples of sequential motor behaviors, share specific motor patterns; rather, they may share similar mechanisms for their implementation as well as adhere to similar organizational principles (see Grillner, 1982; Kelso & Tuller, 1984). Common organizational principles and sensorimotor processes may be used for speech and other motor behaviors, although they will be adopted to specific task requirements (e.g., communication) and effector properties. Speech and other sequential motor behaviors such as typing, handwriting, locomotion, mastication, and to a lesser extent respiration involve serial ordering of muscle actions and movements. For more automatic behaviors such as mastication and locomotion, central rhythm generators have been identified which produce behavior-specific rhythmic motor output similar in form and function to those identified in lower vertebrates. Differences in muscle activity and movement patterns for speech, chewing, and respiration clearly indicate that the same central pattern generator does not underlie all behaviors (Moore, Smith, & Ringel, 1988; Smith & Denny, 1990).

A number of observations, however, are consistent with the presence of some kind of rhythm generating mechanism or neural network as the basis for sequential speech motor adjustments. For example, compensatory adjustments for lower lip perturbations during an oral closing movement demonstrate changes in interarticulator timing consistent with the operation of an underlying oscillatory or rhythm generating mechanism (Gracco & Abbs, 1988; 1989). Specifically, the timing of the oral closing action is advanced (vowel duration is shortened) if the perturbation occurs prior to the onset of the closing action (Gracco & Abbs, 1988). In a complementary investigation it was also found that if a lip perturbation was unexpectedly removed well in advance of oral closure, the closing action was delayed (vowel duration increased) (Gracco & Abbs, 1989). These results are consistent with a conclusion that phase-related effects of sensory stimuli, resulting from the perturbation, interacting with rhythmic motor output to modify sequential timing. The qualitative observation of spatiotemporal consistency of sequential movements associated with repeated production of sentence-length material (see Gracco, 1990) is also suggestive on an underlying sequencing mechanism. Other results such as minimal movement durational changes to static (Lindblom, Lubker,

Gay, Lyberg, Branderal, & Holgren, 1987) and dynamic perturbation (Gracco & Abbs, 1988) are consistent with an underlying mechanism in which sequential timing is maintained.

Recent experiments and theoretical perspectives on the neural control of rhythmic respiratory movements offer an interesting framework for speech movement sequencing (Feldman, Smith, McCrimmon, Ellenberger, & Speck, 1988). It has been suggested that the central pattern generator for respiration may more appropriately be regarded as two separate, but interacting, processes; one specifying the pattern of muscle actions, and one specifying the timing of the output (the rhythm). A similar scheme can be suggested for speech. The characteristic neuromotor patterns for speech sounds outlined above interact with a central rhythm generating process which dictates the timing of the output (see also Saltzman & Munhall, 1989). Two studies of note have attempted to evaluate the apparent rhythmicity of speech. Ohala (1975) recorded over 10,000 jaw movements over a 1.5 hour period of oral reading. Although there were frequencies evident from spectral analysis in the range of 2-6 Hz significant variability was also observed. In contrast, Kelso et al. (1985) reported a rather strong periodicity, with little variability, at approximately 5-6 Hz for lower lip/jaw movements during reiterant speech. The results of the two studies are only contradictory if one assumes that context should not interactively affect rhythmic output. The Ohala study did not constrain the reading material and, hence, reflected a range of phonemic content. Kelso and colleagues, on the other hand, restricted the phonemic content to "ma" and "ba" It seems more likely, given the intrinsic timing character of various sounds, that output frequency may be modulated by phonemic context; the sounds of the language may have their own intrinsic frequency (timing) properties (cf. Fowler, 1980). For example, vowels can be categorized as long or short, generally related to their average relative duration, and consequently to different speed and extent of jaw opening actions. Similarly, movements of various articulators associated with high pressure consonants are often produced at a faster rate than their voiced low pressure counterparts. As shown recently, the oral closing movement is initiated sooner with a tendency for higher closing movement velocity when the consonant is /p/ as opposed to /b/ or /m/ (Gracco, submitted). It is suggested that a central rhythm generator provides the framework for the sequencing of sound-specific patterns with contain certain intrinsic phoneme-specific differences resulting in the continuous modulation of the basic rhythm.

An important consequence of incorporating a central rhythm generator into a speech production model is the ability to explain rate, stress, and final lengthening changes with manipulation of a single mechanism; global and local changes in the frequency of the rhythm. Changes in speaking rate can be viewed as an increase in the output of the generator, producing characteristic changes in the segments as well as their sequencing. For example, increasing the output frequency of the generator (increasing speech rate) is accompanied by higher amplitude, shorter duration bursts of muscle activity (see Figure 4 for example, also Gay, Ushijima, Hirose, & Cooper, 1974; Gay & Hirose, 1973) which results in higher movement velocities, as shown in Figure 4, and a reduction in movement displacement (Kelso et al., 1985). The reduction in movement displacement is a consequence of greater gestural overlap (Browman & Goldstein, 1989; Saltzman & Munhall, 1989) effectively increasing the damping. Similarly, stress and final lengthening can be viewed as a local decrease in the output frequency. It is the case that phrase-final lengthening and stress manifest different kinematic effects (see Edwards, Beckman, & Fletcher, 1991). However, these may merely reflect differences in context such that phrase final articulations are less constrained because of the relative time between it and the next segment, and the movement continues longer and farther as a consequence; there is no active mechanism to arrest the movement. The possibility that a central rhythm generator underlies the serial timing is an attractive hypothesis that is in need of empirical validation.

POTENTIAL NEURAL MECHANISMS

From the previous discussion, it has been suggested that there are multiple functional processes underlying the generation and sequencing of speech movements. These processes include phonological (vocal tract) specification, sensorimotor integration, and sequencing of sound-producing elements. A fundamental premise in the present model is that there are characteristic patterns stored in the nervous system whose selection and activation initiate events which ultimately produce coordinated sequential vocal tract actions. At present any attempt to speculate on where or how such patterns are stored would be premature.

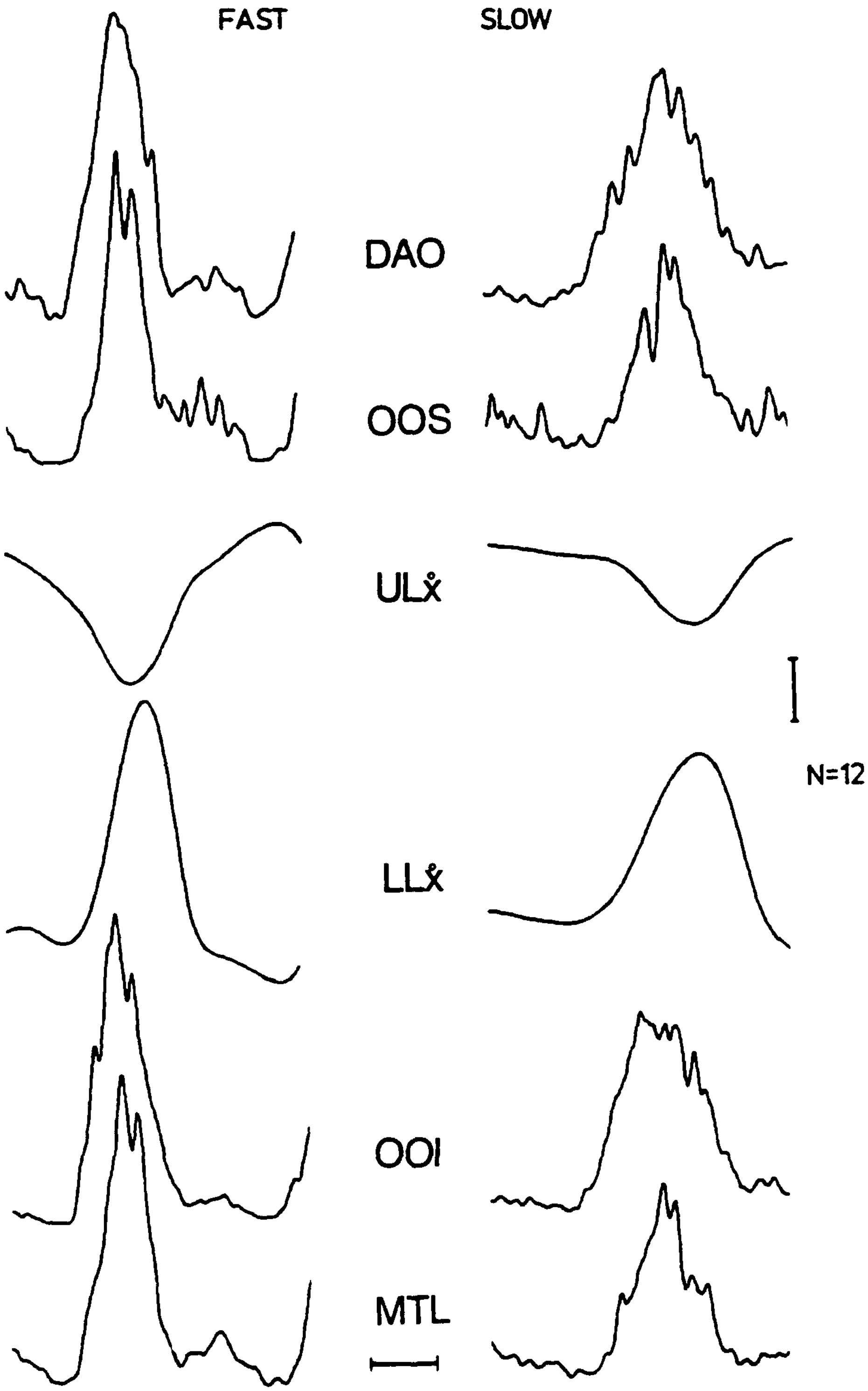

Figure 4. **Averaged (n=12) muscle activity for upper lip and lower lip muscles and the associated upper and lower lip closing movement velocities. Subject repeated the word "sapapple" at a fast and slow (subject defined) rate. Averages were aligned to the peak jaw opening velocity (not shown). Although the peak velocities are higher during the fast rate condition, compared to the slow rate condition, the resulting displacements are smaller.**

However, it is possible to consider the sensorimotor implementation of these hypothetical patterns as well as to generally speculate on the contribution of various distributed neuroanatomical systems that are known to be involved in speech production (cf. Abbs, 1986; Gracco & Abbs, 1987; Kent, 1990 for reviews).

In humans, acquired lesions posterior to the central sulcus result in a form of fluent aphasia characterized by varying degrees of phonological impairment (Blumstein, Cooper, Zurif, & Caramazza, 1977; Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Tuller, 1984). Given the large representation of facial structures, and the projections to supplementary and premotor cortices (Petrides & Pandya, 1984; Wiesendanger & Weiesendanger, 1984), posterior parietal cortex (area 7b), having sensory, motor, and behavioral functions (Hyvärinen, 1981; 1982), seems a likely candidate for the instantiation of phonological goals. As suggested above, it is not clear where the phonological specifications are stored, but once recalled from memory the posterior parietal region may be involved in the setting up of a number of neuroanatomical system used for the implementation of speech motor actions. As such, posterior parietal and no doubt portions of frontal cortex, are "upstream" from the sensorimotor implementation of speech production and can be viewed as performing a prescriptive or executive function.

In contrast, two major brain systems, involving the basal ganglia and supplementary motor area (SMA) and the cerebellum and pre-motor area (PM), are viewed as the major implementation centers to carry out the details of the speech production process. The function of the basal ganglia-SMA system, surmised from human lesion and behaving nonhuman primate studies, appears to have the requisite function to be involved in scaling the hypothesized characteristic neuromotor patterns in the present model. For example, behavioral data from the human limb studies (see Marsden, 1984 for review) and focal stimulation and lesion data from behaving nonhuman primates in which the primary deficit was an inability to scale muscle actions (DeLong, Alexander, Georgopoulos, Crutcher, Mitchell, & Richardson, 1984; Horak & Anderson, 1984a,b). SMA lesions appear to exaggerate the inability to scale muscle actions to task, often resulting in total speech arrest (Arseni & Botez, 1961; Caplan & Zervas, 1978) and a pronounced reduction in self-initiated voluntary movement (see Wiesendanger, 1985 for review). Parkinson's disease results in speech movement impairments that reflect generalized reduction in the speed, and extent of articulatory movements resulting in perceptually distorted consonants, slowed speech rate, and a tendency toward monotone. It is suggested that these deficits reflect a generalized reduction in the ability to scale muscle actions to the specific speech movement requirements. Consistent with the location of the basal ganglia upstream from motor cortex and the relatively indirect access of direct sensory information, it is suggested that the neuromuscular scaling operation is controlled by cortical influence, predominantly the SMA with secondary influences from other cortical areas (Alexander, DeLong, & Strick, 1986).

Speech movement deficits associated with Parkinson's disease do not demonstrate impairments in the duration of the individual movements (Connor, Abbs, Cole, & Gracco, 1989: Forrest et al., 1989) suggesting that the basal ganglia is not involved in the sequencing of movements. However, aphasic patients with anterior cortical lesions and ataxic dysarthrics demonstrate a sequencing difficulty manifest in voice onset timing (see Baum, Blumstein, Naeser, & Palumbo, 1990; Blumstein et al., 1977; 1980), a sequencing difficulty consistent with damage to the premotor area which receives projections from the cerebellum, a neural structure involved in timing movement sequences (Kent & Rosenbeck, 1982; Gracco & Abbs, 1987; Ito, 1984). Similarly, neurophysiological investigations in nonhuman primates have shown the PMA to be involved in the sensory guidance of movements (Godschalk, Lemon, Nijs, & Kuypers, 1981; Halsband & Passingham, 1982; Rizzolatti, Scandolaara, Matelli, & Gentilucci, 1981) similar to the function proposed for the cerebellum (Ito, 1984; Soechting, Ranish, Palminteri, & Terzuolo, 1976). In general, the cerebellar-PM system appears to function as an important component in the incorporation of peripheral sensory signals into the central motor commands.

The final component in the present model is the hypothesized central rhythm generator. While there is no evidence that the cerebellum is the site of a central rhythm generator for any motor action, it has been suggested by Ito (1984) that the cerebellum may contribute to the timing of many rhythmic motor behaviors. The speech timing changes associated with cerebellar damage is consistent with at least a contributing role. Other considerations for the locus of a central rhythm generator would be the intricate synaptic

connections within the brainstem that could possibly be temporarily set into oscillation by directed input from cortical structures, similar to the central masticatory rhythm generator (Nakamura, 1986 for example). An alternate possibility is that speech rhythm and hence serial timing is a network property that emerge from a hierarchical organization (Martin, 1972). It is clear that a definitive answer to the presence and possible location of a central rhythm generator underlying speech timing will require a great deal more experimental consideration.

One prediction from the sensorimotor organization presented in the present chapter in which the vocal tract is considered the smallest functional control structure operated on by sensorimotor scaling and timing processes is the absence of subphonemic speech errors as would occur with speech subsystem impairment (Abbs, Hunker, & Barlow, 1983). Except for cases of focal nervous system damage such as a dystonia, or lower motoneuron damage, speech motor impairments specific to an articulatory subsystem should not occur. The deficits associated with various nervous system damage may result in different degrees of impairment because of the biomechanical or physiological differences of individual articulators. However, it is not clear that surface differences are a true reflection of underlying differential deficits. For a variety of speech motor disorders due to damage to basal ganglia, cerebellum and anterior and posterior cortical areas, deficits are observed that are consistent with a global rather than focal breakdown. That is, the major neuroanatomic sensorimotor systems involved in speech production including the basal ganglia-supplementary motor system, cerebellar-premotor cortical system, and inferior parietal cortex, appear to function, not in the control of movement per se, but in processes from which movement emerges.

SUMMARY

The framework that emerges from the preceding is that speech motor control involves a small number of sensorimotor processes applied in a unitary manner to the vocal tract and modulated according to task requirements such as speech rate, articulatory precision, and suprasegmental stress. In the current model, these processes include selection and activation of characteristic vocal tract actions, spatiotemporally scaled according to phonological considerations, such as intrinsic timing properties, and peripheral conditions. Somatic sensory information is an important component of the system allowing dynamic modulation of relatively stereotypic motor commands. An underlying rhythmic mechanism is proposed which provides the temporal framework for sequential speech adjustments as well as a mechanism to systematically vary suprasegmental speech timing. These fundamental sensorimotor processes interact and overlap to produce the continuous dynamic modulation of the vocal tract generating time-varying pressures and flows. An important constraint on the model is that the underlying processes are consistent with generally accepted nervous system operations. An important prediction from the model is that nervous system damage, unless extremely focal, should produce global deficits attributable to one or some combination of three major nervous system functions for speech; pattern specification, scaling of muscle actions, and initiation and sequencing of the production units.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscles responses to load perturbation of the lip during speech. *Journal of Neurophysiology, 51,* 705-723.

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior, 16,* 195-232.

Abbs, J. H., Hunker, C., & Barlow, S. (1983). Differential speech motor subsystem impairments with suprabulbar lesions: Neurophysiological framework and supporting data. In W. Berry (Ed.), *Clinical dysarthria* (pp. 21-56). College Hill Press: San Diego.

Atkeson, C. G., & Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience, 5,* 2318-2330.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357-381.

Arseni, C., & Botez, M. I. (1961). Speech disturbances caused by tumors of the supplementary motor area. Acta Psychiatrica Neurologica Scandinavica, 36, 279-298.

Barlow, S. M., & Abbs, J. H. (1978). Some evidence of auditory feedback contributions to ongoing speech motor control. Paper presented at the American Speech and Hearing Convention, San Francisco, CA.

Baum, S., Blumstein, S., Naeser, M., & Palumbo, C. (1990). Temporal dimensions of consonant and vowel production: an acoustic and CT scan analysis of aphasic speech. *Brain and Language, 39,* 33-56.

Black, J. W. (1951). The effect of delayed sidetone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders, 16,* 56-60.

Blumstein, S., Cooper, W., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: a voice-onset time analysis. *Brain and Language, 9,* 153-170.

Blumstein, S., Cooper, W., Zurif, E., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15,* 371-383.

Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain and Language, 7,* 309-319.

Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). Academic Press: New York, .

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 250-251.

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18*, 299-320.

Caplan, L. R., & Zervas, N. T. (1978). Speech arrest in a dextral with a right mesial frontal astrocytoma. *Archives of Neurology, 35*, 252-253.

Connor, N. P., Abbs, J., Cole, K. J., & Gracco, V. L. (1989). Parkinsonian deficits in serial multiarticulate movements for speech. *Brain, 112*, 997-1009.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. The MIT Press: Cambridge, Mass.

Churchland, P. M. (1989). A neurocomputational perspective: The nature of mind and the structure of science. MIT Press: Cambridge, Mass.

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics, 1*, 239-248.

DeLong, M. R., Alexander, G. E., Georgopoulos, A. P., Crutcher, M. D., Mitchell, S. J., & Richardson, R. T. (1984). Role of basal ganglia in limb movements. *Human Neurobiology, 2*, 235-244.

Edelman, S., & Flash, T. (1987). A model of handwriting. *Biology of Cybernetics, 57*, 25-36.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of Acoustical Society of America, 89*, 369-382.

English, A. W. (1982). Are whole muscles the fundamental substrate for the CNS control of movement? *Behavior and Brain Sciences, 5*, 544-545.

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders, 20*, 333-346.

Feldman, J. L., Smith, J. C., McCrimmon, D. R., Ellenberger, H. H., & Speck, D. F. (1988). Generation of respiratory patterns in mammals. In A. H. Cohen, S. Rossignol, & S. Grillner (Eds.), *Neural control of rhythmic movements in vertebrates* (pp. 73-100). New York: Wiley and Sons.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech Hearing Research, 18*, 207-220.

Folkins, J. W., & Linville, R. N. (1983). The effects of varying lower-lip displacement on upper-lip movements: Implications for the coordination of speech movements. *Journal of Speech and Hearing Research, 26*, 209-217.

Forrest, K., Abbas, P., & Zimmermann, G. N. (1986). Effects of white noise masking and low pass filtering on speech kinematics. *Journal of Speech and Hearing Research, 29*, 549-562.

Forrest, K., Weismer, G., & Turner, G. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkisonian and geriatric adults. *Journal of the Acoustical Society of America, 85*, 2608-2622.

Fowler, C. A., (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics 8*, 113-133.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: Human Perception and Performance, 112*, 386-412.

Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth, (Ed.), *Language production* (pp. 373-420). New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Journal of Phonetics, 37*, 306-326.

Fowler, C. A., Gracco, V. L., & V.-Bateson, E. (1989). Remote and local effects of stress within and among articulatory subsystems. *Journal of the Acoustical Society of America, 86(1)*, S115.

Flashner, H., Beuter, A., & Arabyan, A. (1988). Fitting mathematical functions to joint kinematics during stepping: Implications for motor control. *Biology of Cybernetics, 58*, 91-99.

Gay, T., & Hirose, H. (1973). Effect of speaking rate on labial consonant production. *Phonetica, 27*, 44-56.

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics, 2*, 47-63.

Godschalk, M., Lemon, R., Nijs, H., & Kuypers, H. (1981). Behavior of neurons in monkey peri-arcuate and precentral cortex before and during visually guided arm and hand movements. *Experimental Brain Research, 44*, 113-116.

Gracco, V. L. (1987). A multilevel control model for speech motor activity. In H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 57-76). Wien: Springer-Verlag.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8*, 4628-4634.

Gracco, V. L. (1990). Characteristics of speech as a motor control system. In G. R. Hammond (Ed.), *Cerebral control of speech and limb movements* (pp. 3-28). North-Holland: Elsevier.

Gracco, V. L. (submitted). Some organizational principles for speech movement control. *Journal of the Acoustical Society of America*.

Gracco, V. L. & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysics, 54*, 418-432.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

GraccoV. L., & Abbs, J. H. (1987). Programming and execution processes of speech motor control. Potential neural correlates. In E. Keller & M. Gopnick, Eds.), *Motor and sensory processes of language* (pp. 163-202). Hillsdale, NJ: Lawrence Erlbaum.

Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research, 71*, 515-526.

Gracco, V. L., & Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research, 75*, 586-598.

Gracco, V. L., & Löfqvist (1989). Speech movement coordination: Oral-laryngeal interactions. *Journal of the Acoustical Society of America, 86*, S114.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblon, J. Lubker, & A Persson (Eds.), *Speech motor control* (pp. 217-230). Oxford: Pergammon Press.

Halsband, U., & Passingham, R. (1982). The role of premotor and parietal cortex in the direction of action. *Brain Research, 240*, 368-372.

Hollerbach, J. M. (1981). An oscillation theory of handwriting. *Biological Cybernetics, 39*, 139-156.

Horak, F. B., & Anderson, M. E. (1984a). Influence of globus pallidus on arm movements in monkeys. I. Effects of kainic acid-induced lesions. *Journal of Neurophysiology, 52*, 290-304.

Horak, F. B., & Anderson, M. E. (1984b). Influence of globus pallidus on arm movements in monkeys. II. Effects of stimulation. *Journal of Neurophysiology, 52*, 305-322.

Hyvärinen, J. (1981). Regional distribution of functions in parietal association area 7 of monkey. *Brain Research, 206*, 287-303.

Hyvärinen, J. (1982). *The parietal cortex of monkey and man*. Berlin: Springer-Verlag.

Ito, M. (1984). *The cerebellum and neural control*. Raven Press: New York.

Jackson, J. Hughlings (1875). *Clinical and physiological researches on the nervous system*. (Contains a reprinting of "On the anatomical and physiological localisation of movement in the brain," first published in Lancet, 1873, i:84-85). Churchill: London.

Joos, M. (1948). Acoustic phonetics. *Language, 24*, 1-136.

Kelso, J. A. S., & Tuller, B. (1983). "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research, 26*, 217-224.

Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamic principles for speech and movement coordination. *American Journal of Physiology, 15*, R928-R935.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Kelso, J. A. S., Tuller, B, V.-Bateson, E., & Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 812-832.

Kelso, J. A. S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 105-128). Berlin: Springer-Verlag.

Kent, R. D. (1990). The acoustic and physiological characteristics of neurologically impaired speech movements. In W. Hardcastle & A Marchal, (Eds.), *Speech production and modelling* (pp. 365-401). Kluwer Academic Publishers: Dordrecht.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5*, 155-133.

Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbances and neurologic lesion. *Brain Language, 15*, 259-291.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach (Ed.), *Tutorials in motor behavior* (pp. 3-47). Amsterdam: Elsevier/North Holland.

Lacquaniti, F. (1989). Central representations of human limb movement as revealed by studies of drawing and handwriting. *Trends in Neuroscience, 12*, 287-291.

Lane, H. L., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14*, 677-709.

Lindblom, B., Lubker, J., Gay, T., Lyberg, P., Branderal, P., & Holgren, K. (1987). The concept of target and speech timing. In R. Channon & L. Shockery (Eds.), *In honor of Ilse Lehiste* (pp. 161-181). Dordrecht, The Netherlands: Foris Publications.

Loeb, G. E. (1985). Motoneurone task groups: Coping with kinematic heterogeneity. *Journal of Experimental Biolology, 115*, 137-146.

Löfqvist, A. (1990). Speech as audible gestures. In W. Hardcastle & A Marchal, (Eds.), *Speech production and modelling* (pp. 289-322). Kluwer Academic Publishers: Dordrecht.

Löfqvist, A., & Yoshioka, H. (1981). Interarticulator programming in obstruent production. *Phonetica, 38*, 21-34.

Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication, 3*, 279-289.

Marr, D. (1982). *Vision*. New York: W. H. Freeman & Company.

Marsden, C. D. (1984). Which motor disorder in Parkinson's disease indicates the true motor function of the basal ganglia? In D. Evered & M. O'Connor, (Eds.), *Functions of the basal ganglia* (pp. 225-236). Pitman: London.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487-509.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol 2: Applications.* MIT Press: Cambridge, MA.

Marr, D. (1982) *Vision*. W. H. Freeman: San Francisco.

Moore, C. A., Smith, A., & Ringel, R. L. (1988). Task-specific organization of activity in human jaw muscles. *Journal of Speech and Hearing Research, 31*, 670-680.

Morasso, P. (1981). Spatial control of arm movements. *Experimental Brain Research, 42*, 223-227.

Morasso, P., & Mussa-Ivaldi, F. A. (1982). Trajectory formation and handwriting: A computational model. *Biological Cybernetics, 45*, 131-142.

Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech *Journal of Phonetics, 20*, 111-126.

Munhall, K., Löfqvist, A., & Kelso, J. A. S. (in press). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America.*

Nakamura, Y. (1986). Neural basis of rhythm generation of masticatory jaw movements. *Progressive Neural Research, 30*, 237-250.

Ohala, J J. (1975). The temporal regulation of speech. In G. Fant & M. A. A. Tatham, (Eds.), *Auditory analysis and perception of speech* (pp. 431-454). London: Academic.

Öhman, S.E.G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Pellionisz, A., & Llinas, R. (1979). Brain modeling by tensor network theory and computer simulation. The cerebellum: Distributed processor for predictive coordination. *Neuroscience, 4*, 323-348.

Pellionisz, A., & Llinas, R. (1985). Tensor network theory of the metaorganization of functional geometries in the central nervous system. *Neuroscience, 16*, 245-273.

Perkell, J. S. (1980). Phonetic features and the physiology of speech production. In B. Butterworth, (Ed.), *Language production* (pp. 337-372. New York: Academic Press.

Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *Journal of Comparative Neurology, 228*, 105-116.

Ringel, R., & Steer, M. D. (1963). Some effects of tactile and auditory alterations on speech output. *Journal of Speech and Hearing Research, 6*, 369-378.

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. I. Somatosensory responses. *Behavioral Brain Research, 2*, 125-146.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing*. Cambridge, Massachusetts: The MIT Press.

Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In, H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 129-144). Berlin: Springer-Verlag, .

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94*, 84-106.

Saltzman, E. L., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333-382.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86*, 78-87.

Smith, A., & Denny, M. (1990). High frequency oscillations as indicators of neural control mechanisms in human respiration, mastication, and speech. *Journal of Neurophysiology*.

Soechting, J. F., Ranish, N. A., Palminteri, R., & Terzuolo, C. A. (1976). Changes in a motor pattern following cerebellar and

olivary lesions in the squirrel monkey. *Brain Research*, 105, 21-44.

Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. North Holland: Amsterdam.

Tuller, B. (1984). On categorizing aphasic speech errors. *Neuropsychologia*, 22, 547-557.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology* (pp. 211-265). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Weisendanger, M. (1985). Recent developments in studies of the supplementary motor area of primates. *Review of Physiological Pharmacology and Biochemistry*, 103, 1-59.

Wiesendanger, M., & Wiesendanger, R. (1984). The supplementary motor area in light of recent investigations. *Experimental Brain Research Supplement*, 9, 382-392.

Zimmermann, G. N., Brown, C, Kelso, J. A. S., Hurtig, R., & Forrest, K. (1988). The association between acoustic and articulatory events in a delayed auditory feedback paradigm. *Journal of Phonetics*, 16, 437-451.

FOOTNOTE

*To appear in H. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering*.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 45-58

Analysis of Speech Movements: Practical Considerations and Clinical Application

Vincent L. Gracco

The instrumental evaluation of speech movements is an important adjunct to the assessment and understanding of speech motor disorders. As the interface between the nervous system and aerodynamic modifications in the vocal tract, movement variables such as displacement, velocity, acceleration, and their time histories, can provide direct information on speech motor disorders that can only be inferred from acoustic or perceptual evaluation. Impairment in various aspects of neuromotor functioning is reflected in the motion of individual articulators and their coordination, and may reflect early signs of functional change due to disease or trauma. Within certain limits, movement analysis can be used as an objective method for categorizing speech motor disorders and monitoring change due to therapeutic intervention. Further, objective comparison of orofacial motor behavior during speech and nonspeech tasks may provide diagnostic insight into underlying pathophysiological processes. A perspective on the potential utility of speech movement analysis in the assessment, treatment, and understanding of speech motor disorders is the focus of the present chapter. The limitations of speech movement analysis and the need for clinically-relevant research will be presented.

INTRODUCTION

With the increased availability of measurement devices for transducing movements of the speech articulators, computer software for automated processing and analysis of data, and decreased cost of computer hardware, instrumental evaluation of human vocal tract movements is becoming more feasible for inclusion into the clinic. Analysis of upper and lower limb movements employing various instrumental tests have been used for the last 40 years to aid in the evaluation and diagnosis of various pathophysiological conditions and to determine the outcome of clinical trials (see Potvin & Tourtellotte, 1985 for review). For speech, movement analysis is a potentially important adjunct to more traditional acoustic and perceptual analyses used routinely in the clinic. In addition, analysis of speech and nonspeech (orofacial) movements can be used to evaluate the consequences of motor disorders that have not yet developed to the point of significantly affecting the communicative process. The purpose of the present chapter is to outline some of the ways in which analysis of movement parameters and movement patterns may be used clinically. Before proceeding, it may be helpful to reiterate a point made by Potvin and Tourtellotte (1985);

> "To the extent that instrumented tests can be developed for measuring functions, their selective use can provide information that might not otherwise be available. However, investigators should be aware that the ability to measure small differences reliably can yield statistically significant differences that may not be of clinical importance."

In the following, the focus will be on measurements that may have specific functional utility in terms of assessing speech production capabilities, detecting differences in neurologic function, and improving understanding of speech motor performance. Because of the current limitation in normative data and the wide range of inter- and intrasubject variability, both qualitative and quantitative methods will be presented.

This work was supported in part by NIDCD Grants DC-00594 and DC-00121.

INTERPRETATION OF MOVEMENT

The evaluation of movement can be approached from a variety of perspectives. From a motor control perspective, speech production is observed to be a sequential production of different vocal tract configurations that are coordinated in space and time and overlap to various degrees. Visual inspection of speech movements allows for a qualitative impression of overall motor functioning. Compare, for example, the lip and jaw movement signals presented in the left half of Figure 1, obtained from a neurological normal subject, with the movement signals in the right half of the figure, obtained from a subject with Parkinson's disease (PD). Each subject is repeating the same sentence and the scaling for the two sets of signals is the same. Without knowing what is being said, and disregarding the respective acoustic signals, it can be seen that there are marked differences in the two sets of movements. While there are some general similarities in the overall movement patterns, the extent of articulator motion of both the upper lip and lower lip/jaw movements for the PD subject is less than for the normal subject, consistent with the clinical manifestations of hypokinesia. Movement velocities, displayed above and below the respective UL and LLJ displacements, are severely reduced in magnitude for the PD subject as well. Further insight can be gained into the manifestations of the disorder by evaluating the acoustic signal simultaneously with the movement signals. The impoverished and slow movements from the Parkinson's subject are accompanied by a poorly differentiated acoustic signal consistent with the perceptual speech characteristics of imprecise consonant production.

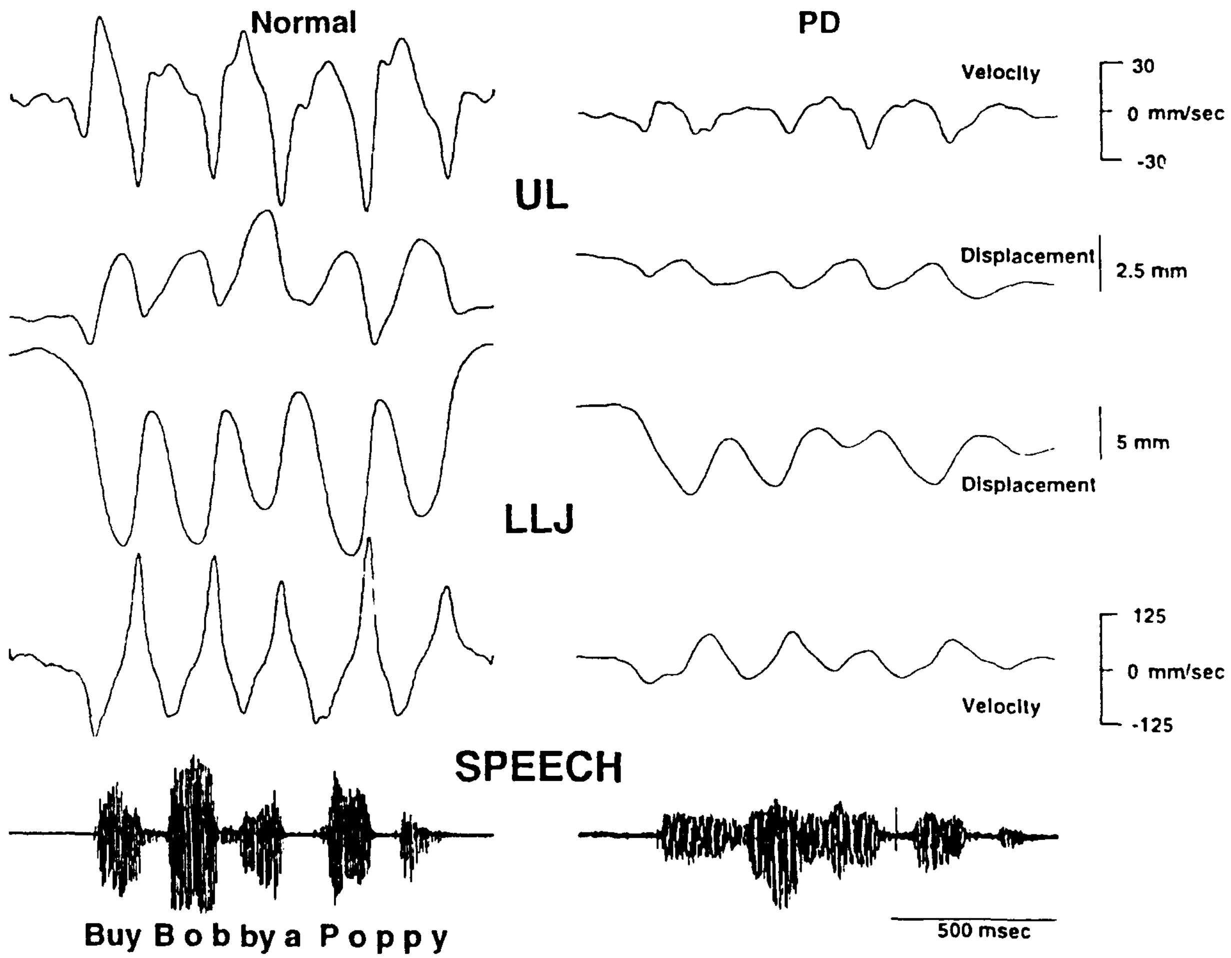

Figure 1. Upper lip (UL) and lower lip/jaw (LLJ) movement displacement and velocity from a neurologically normal subject and a subject with Parkinson's disease (PD). The subjects task was to repeat the utterance "Buy bobby a poppy" at a comfortable rate and loudness with even stress. Shown below each set of movement signals is the respective acoustic speech signal.

Other qualitative observations can be made from movement signals that are important for a thorough understanding of the sensorimotor breakdown and functional deficits associated with particular speech disorders. Based on previous research it has been shown that multiple articulators engaged in the production of the same sound display spatial and temporal patterns that reflect their cooperative behavior (Gracco, 1988, 1990; Gracco & Löfqvist, 1989). Individual speech movements generally display smooth continuous motion characterized by a unimodal velocity profile (Gracco & Abbs, 1986; Munhall, Ostry, & Parush, 1985; Nelson, 1983; Ostry, Cooke, & Munhall, 1987). Breakdown in the coordinative action of multiple articulators, a loss in the ability to smoothly sequence concatenated vocal tract gestures, or multiple peaks in the velocity profile associated with a single articulatory movement are observations that reflect qualitatively on the processes of speech motor control. From examination of discrete events associated with a single speech or nonspeech motor task, it is also possible to functionally evaluate the neuromotor system at the level that reflects on the net force applied to articulators to produce individual movements. In order to generate movement a certain pattern of excitation and inhibition is produced in the nervous system and directed to the lower motor neurons. The action potentials generated by the input signals result in two distinct peripheral events; electrical responses in the muscle membranes producing EMG's, and the generation of forces originating from the contractile elements of the muscles. Movement reflects the summation of net active and passive forces with a certain time history filtered through the biomechanical properties of the structures being moved. If the structure is at least in part inertial, the initial acceleration of the load will be proportional to the initial contractile force. Similarly, the peak velocity of a movement is generally proportional to the force magnitude integrated over the movement time. Inspection of individual movement patterns can provide heuristic information regarding the neuromotor functioning of the patient and reflect on the mechanical characteristics of particular articulators.

BASIC KINEMATICS

In order to objectively and quantitatively evaluate speech movements a measurement framework is required. Any description of movement relies on the terminology of kinematics. A complete kinematic description of any movement, especially of the vocal tract, is geometrically complex. For most purposes, the motion of bodies can be reduced from irregular shaped masses to points, and the motion of such points can be described with kinematic variables. The description of point motion is analytically complex, requiring 15 data variables which change over time (Winter, 1979). For clinical purposes, the displacement (the distance from a starting to an ending position) and velocity (the directional speed) are the most useful for describing articulatory motion. In order to keep track of the changing kinematic variables and maximize their descriptive usefulness it is important to adopt a reference convention and a coordinate system. Motion can be described relative to some static articulatory position, such as lip movement relative to a rest position. An alternative that also provides spatial information is to reference the movements to an immobile anatomical structure. The most frequently used spatial coordinate system involves three perpendicular axes representing the sagittal, frontal, and transverse planes. Movements of articulators can then be described with respect to inferior-superior (y), anterior-posterior (x), and lateral-medial (z) directions, respectively, relative to some anatomical reference. The most important consideration for clinical use is that a convention be established, one that is consistent with respect to the purpose of the measurement and reproducible within and across subjects.

As mentioned, the displacement of a point on an articulator surface and the velocity at which the articulator moves are two important kinematic variables fundamental to the description and evaluation of motor disorders characterized by hypokinesia (reduction in movement extent), bradykinesia (slowness in movement; reduced velocity), and akinesia (slowness in movement initiation). Shown on the left in Figure 2 is a position time history of a single midsagittal point on the lower lip as it moves from opening for a vowel to oral closure for /p/. Under the displacement signal is the time history of the instantaneous velocity mathematically derived from the displacement signal. From the displayed signals, the maximum displacement, calculated as the distance between onset position and offset position associated with the movement, and the associated peak instantaneous velocity, are easily obtained. Additionally, the duration of the movement, defined as the time from onset to completion can also be obtained. As shown in the figure, the velocity profile can be further dissected to provide information on the accelerative and decelerative phases of the movement.

Ignoring gravity, the accelerative phase of a movement generally reflects the increase in net force applied to the load (articulator) due to the contraction of the muscles. In contrast, the decelerative phase of a movement generally reflects the decrease in net force acting on the load due to the relaxation of the contractile process and any antagonistic muscle actions. Displacement and velocity measures provide the means to describe and quantify movement and also allow some inference on the properties of the muscular actions that caused the motion. In addition to measuring the discrete components of a movement, the frequency and amplitude of repeated productions can also be calculated as illustrated on the right side of Figure 2.

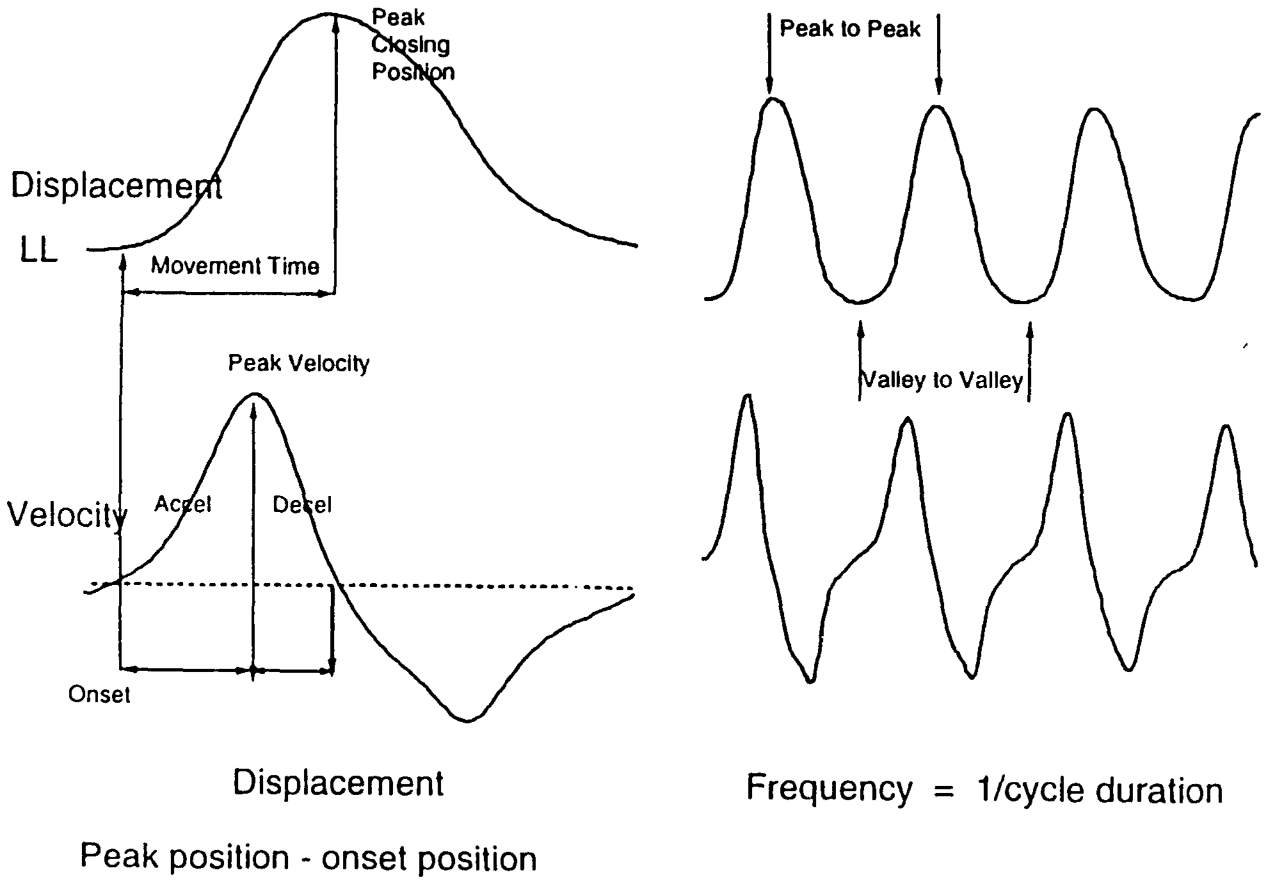

Figure 2. **Representation of the displacement and velocity of a point on the lower lip associated with a single oral closing movement for /p/ (left hand portion of the figure). Shown are some of the variables to be measured (see text for further details). The displacement and velocity of the same point on the lower lip during repetitive opening and closing movements associated with repetition of /pae/. From repetitive syllables, the frequency of production can be derived as shown.**

INSTRUMENTATION

Prior to presenting a protocol that we have been using to instrumentally evaluate speech and non-speech movements, a brief discussion of the movement transduction devices and general operating principles follows. Monitoring upper articulator movement can be accomplished using a variety of transduction techniques. In general, these techniques convert mechanical energy, represented as movement of an articulator or group of articulators, to electrical energy, represented as an analog voltage. Many methods are available to convert a physiological event to an electrical signal and generally involve direct or indirect variation in electrical quantities such as resistance, capacitance, inductance, or the magnetic linkage between coils. The four basic techniques currently available in different forms for use in the speech clinic involve strain gauge transduction, optical transduction (optoelectronic sensing devices), imaging (ultrasound), and electromagnetic transduction. The following will briefly review the techniques and commercially available devices with respect to their basic principles of operation, clinical utility, and practical limitations. A more detailed analysis can be obtained from various sources such as Abbs and Watkin (1976), Baken (1987), and Geddes and Baker (1968).

Strain gauge transduction

Strain gauges are resistive elements that are mounted on a flexible, lightweight strip of metal anchored at one end and attached to a moving surface on the other end. The voltage output from a gauge is proportional to the movement at the end of the mobile attachment. Strain gauge transducers are used for monitoring external articulatory movements such as the lips and jaw. Initially, the technique was used in the transduction of jaw and lip movements by Sussman and Smith (1970a, b). Refinements of the method of attachment have been reported by Abbs and Gilbert (1973) and Müller and Abbs (1979). A significant clinical development was reported by Barlow, Cole, and Abbs (1983) in which strain gauge transducers were attached to a lightweight aluminum frame which could be mounted to a subjects head. This refinement allowed the monitoring of lip and jaw movement without requiring stabilization of the subjects' head; for many neurological patients, head stabilization is an unacceptable condition. The cantilever beams can be instrumented to sense motion in one or two (orthogonal) dimensions, although the two dimensional units and their attachments add significantly to the overall weight and can decrease stability. The cantilever beams are commonly attached to a point on the midsagittal plane (midpoint of the lips and chin) providing inferior-superior and anterior-posterior motion sensing. Strain gauge transducers provide a continuous analog output that can faithfully reproduce the fastest lip and jaw movements. A bridge amplifier is required for each direction of movement to supply an excitation voltage to the resistive elements and to amplify the signal prior to storage or analog-to-digital (A/D) conversion.

Optical transduction

The most notable optical technique for tracking human movement involves a position sensing device and pulsed light-emitting diodes to track points in a two or three dimensional coordinate system (Watsmart, Northern Digital, Inc., of Waterloo, Ontario, Canada; Selspot, Selective Electronics, Inc., of Sweden). Devices that rely on the sensing of LED's are limited in a similar manner to the strain gauge devices in that they can only be used to monitor the external articulators such as the lips and jaw. There are some photoelectric devices that rely on the sensing of light reflection which can be used to monitor tongue movement (Chuang & Wang, 1978; Fletcher, 1982). However, such optical scanning systems for tongue motion require small LED light sources and photosensitive detectors arranged in an artificial palate worn by the patient. In addition to this practical limitation and the lack of commercial availability, a distance dependent error has been reported requiring a refinement in calibration procedures (McCutcheon, Lakshminarayanan, & Fletcher, 1990). A final device, using charge coupled device (CCD) sensors eliminating reflection errors, is currently being marketed (Optotrak, Northern Digital, Inc.). Similar to the optoelectric devices, the CCD device provides three dimensional information on the movement of visible sensors with 0.1 mm accuracy over a one cubic meter volume. These commercial devices can also be purchased with customized software for analog-to-digital conversion, signal processing and automated analysis. The most significant drawback to these systems is the cost which may be as high as $50,000 to $60,000 for a complete three dimensional acquisition and analysis system.

Imaging

The most common imaging device having potential clinical application is ultrasound (see Sonies, 1982 for review). An ultrasound signal is

passed into the body and the differential tissue properties associated with different structural layers provide different reflections to the generated sound. The ultrasound reflections are a series of echoes that can then be detected by the transducer. The longer the echoes take to be reflected, the further the tissue is away from the source. Through a knowledge of the anatomy and the different transmission times, the structures within the path of the ultrasound can be reconstructed. For the human vocal tract, ultrasound can be used to visualize and track motion of soft tissue structures such as the tongue and vocal folds. A number of research studies have employed ultrasound to evaluate the shape and motion of the tongue (Sonies, Shawker, Hall, & Gerber, 1981; Stone, Morish, Sonies, & Shawker, 1987; Stone, Shawker, Talbot, & Rich, 1988), the movement of the tongue dorsum during speech (Keller & Ostry, 1983), movement of the vocal folds during devoicing (Munhall & Ostry, 1985) and tongue motion during swallowing (Stone & Shawker, 1986). While ultrasound devices are commercially available they are costly and are often not optimized for vocal tract use.

Electromagnetic transduction

Using alternating magnetic fields it is possible to track point movement of small transducers placed on the tongue, lips, velum, and jaw in the midsagittal plane. The basic device employs a sinusoidal signal driving a transmitter coil which produces lines of magnetic flux. Small receiver coils, or transducers, moving through the magnetic field are induced with a signal that is proportional to the effective cross-sectional area of the receiver coil and the flux density. If the transmitter and receiver axes are parallel, the magnitude of the induced signal is a measure of the distance between the transmitter and receiver. Recently, a commercially available electromagnetic system for tracking movements of the upper articulators has been developed and marketed under the name of the Articulograph AG100 (Carstens Medizinelektronik, Göttingen, West Germany). This system allows the tracking of up to five small receiver coils placed on various supraglottal articulatory structures in the midsagittal plane. The transmitter assembly is placed on the subjects head and secured in a manner similar to the head mounted movement system developed by Barlow et al. (1983). Although the system is commercially available, development and refinement is continuing (see Tuller, Shao, and Kelso, 1990 for initial evaluation of system performance). The system requires a microcomputer to calculate the x-y positions of each transducer in real time and stores the data on the computer disk. Software routines are provided for data display and analysis. Cost of the system, including a microcomputer, is approximately $42,000. Other magnetic devices are commercially available to record positions and movements of the mandible and the interested reader is referred to an article by Michler, Bakke, and Møller (1987) for further information.

There are a variety of commercial devices for the transduction of speech movements, each with certain strengths and weaknesses. The optoelectric devices are capable of three dimensional motion tracking and provide sophisticated software for analysis; the major limitation is the cost. The headmounted movement system is a low cost alternative that can be used with children and adults. The system can be configured to allow transduction in two dimensions although some problems may arise due to the extra weight of the transducer unit. Ultrasound and the Articulograph are the only devices available that allow transduction of tongue movements. Similar to the optoelectric devices, the cost of the respective equipment is high. For all devices, a certain amount of technical sophistication and a basic understanding of the operating principles is required. A final consideration is the transduction of lower lip and jaw movement. The movement transduced at the lower lip is actually a combination of lower lip and jaw movement. In order to evaluate the separate lower lip and jaw actions during speech or nonspeech movements, both the jaw and lower lip and jaw movements are acquired. The jaw signal is then subtracted from the lower lip/jaw signal yielding net lower lip movement. Using the magnetic device, a transducer coil placed on the midpoint between the lower central incisors, can be used as a reflection of "true" jaw motion. For the optical devices, a custom fitted jaw splint can be used with an additional light emitting diode used to track jaw motion. While it is possible to obtain jaw movement from a sensing device placed on the chin, such placement may result in skin movement artifact (see Kuehn, Reich, & Jordan, 1980). For most clinical applications, the combined movement of the lower lip and jaw may suffice, eliminating the need to factor out the contributions of the two articulators.

Other considerations

Once obtained, the data must be stored in some form for analysis. The storage device may be an

oscillographic recorder with a paper medium, an FM tape recorder, or the signals may be digitized directly to computer disk. Data converted from analog to digital form requires anti-aliasing filtering prior to conversion. The general function of anti-aliasing is to insure that false frequencies not present in the original signal are not introduced into the digitized signal. In order to avoid aliasing, the analog signal must be filtered and then digitized at a rate that is at least twice the cutoff frequency of the anti-aliasing filter. The minimum sampling rate is known as the Nyquist rate and is calculated by doubling the highest frequency contained in the signal of interest. Since speech movements contain mostly low frequencies (generally below 15 Hz), the Nyquist rate could be as low as 30 Hz with the anti-aliasing (low pass) filter having a cut off frequency at 15 Hz. However, a 30 Hz sampling rate provides a poor quality time display with a point sampled only every 33.3 ms. (A movement that lasts approximately 120 ms would be represented by only 4 points.) In order to improve the temporal quality, also important when deriving the velocity of the movement, higher sampling rates are often used. An additional consideration is that hardware filters create phase delays in the signal which vary as a function of the cut off frequency. Therefore, it is generally desirable to use an anti aliasing filter with as high a cut off frequency as possible. Once digitized, the movement signals may be further smoothed in software to eliminate any noise in the signal. Using digital filters time delays can be eliminated and the signal can be filtered at a much lower frequency. Similarly, software differentiation (central difference algorithm) is the preferred method of obtaining first and second derivatives since it does not introduce time distortions to the signal.

MOVEMENT ANALYSIS

Most movement disorders result in a reduction in movement extent (hypokinesia), speed (bradykinesia), a slowness in initiation (akinesia), or become generally dyscoordinated. Each of these clinical signs can be evaluated kinematically and subsequently quantified for intrasubject comparisons. We have recently been using a limited speech and oral motor inventory with subjects having various movement disorders focusing on movements of the lips and jaw. Subjects are requested to produce syllables and nonspeech gestures at two rates; a comfortable (preferred) and maximal rate. Words and sentences are also repeated at a comfortable rate and are used for both qualitative and quantitative examination. Nonspeech movements are used to evaluate the orofacial motor system to determine the extent of neuromuscular involvement. It is felt that this protocol provides the minimal amount of information necessary to understand the functional and structural changes accompanying many motor disorders. In the following, movement data for a portion of the protocol will be presented from two subjects, both with PD, who have different degrees of speech motor impairment. Subject one (S1) has minimal speech motor involvement while subject two (S2) has a moderately severe dysarthria characterized by imprecise consonants. Motion of the upper lip and lower lip/jaw were transduced using a head mounted movement system (Barlow et al., 1983) instrumented with strain gauges aligned for two dimensional sensing. The head mounted frame was oriented such that inferior-superior and anterior-posterior movements were referenced to the Frankfort plane.

An initial step in the analysis involves examination of some of the data in two dimensional space. Shown in Figure 3 is the path of the jaw in x-y space, with anterior-posterior movements represented on the x axis and inferior-superior movements represented on the y axis, for a series of speech and nonspeech opening and closing movements. The subject produced repetitive opening and closing movements of the lip/jaw and repeated the syllable /sa/ for approximately 5 seconds the two rates; a comfortable (preferred) and fast (maximal) rate. A number of observations can be made from the x-y representation. First, the increase in speed required for the fast rates results in a general reduction in the movement extent for each task. Second, the extent of movement for /sa/ repetitions is less than that for opening and closing the mouth and the /sa/ repetitions are produced in the middle two thirds of the space occupied by the opening and closing nonspeech movements. Finally, the path taken by the jaw in both tasks and conditions is essentially straight and smooth. These observations from an individual with Parkinson's disease are qualitatively similar to those made for normal subjects. Figure 4, in contrast, displays similar data obtained from S2. As mentioned, this subjects' speech motor skills are more severely affected than the previous subject. From the x-y representations of the speech and nonspeech movements it can be seen that the lip/jaw movements are reduced in extent, less smooth, and more variable than was observed in the previous figure (note the different scales for the two figures).

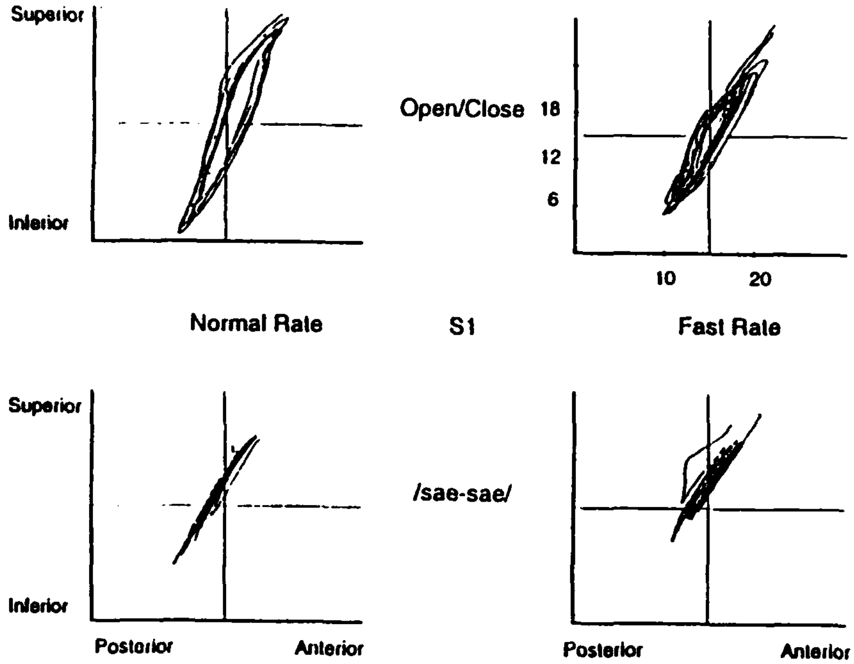

Figure 3. **Two dimensional movement of the lower lip/jaw for repetitive productions of oral opening/closing (nonspeech) and /sae/ for S1 (see text). Movement directions as indicated.**

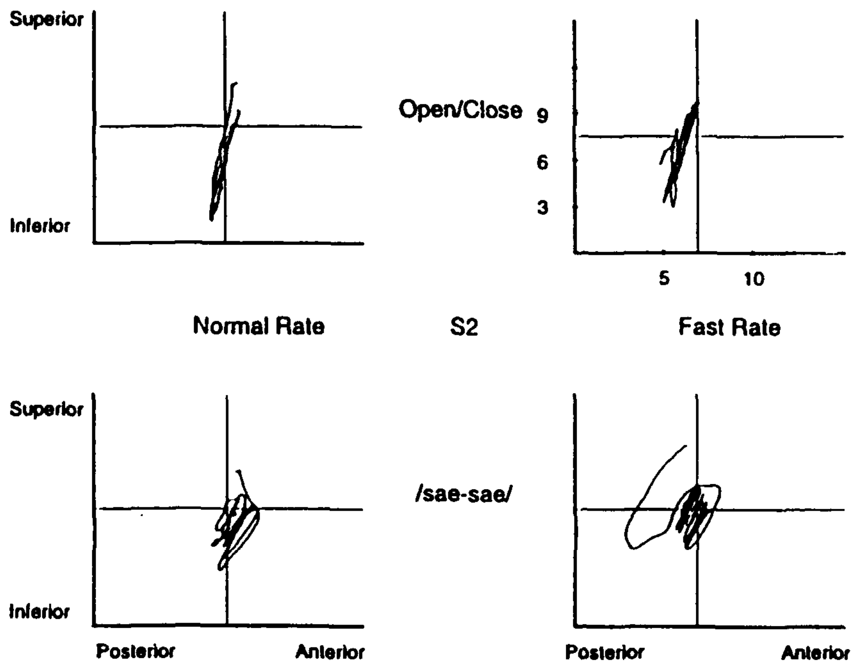

Figure 4. **Two dimensional movement of the lower lip/jaw for repetitive productions of oral opening/closing (nonspeech) and /sae/ for S2 (see text). Movement directions as indicated.**

In order to evaluate such data quantitatively the time histories of the movements must be displayed. Presented in Figure 5 are examples from the two subjects of continuous opening and closing inferior-superior movements (nonspeech) of the LLJ. The well defined peaks and valleys in the displacement trace provides a way of automatically identifying the different movement phases (opening-closing) and calculating the displacement and frequency of repetition. Below each trace is the summary of a software routine which identifies the peaks and valleys in the displacement trace and calculates the frequency of repetition (F0), and the average displacement (mm) of the sequential movements.

Shown in Figure 6 are the upper lip and lower lip/jaw movements in the x and y dimensions associated with repeated production of the syllable /pae/. It can be seen that the upper and lower lips move in both a superior-inferior and anterior-posterior direction. The movements are generally smooth and regular, and the upper lip moves less in extent than the lower lip. Shown in the next figure (Figure 7) are examples from the two subjects illustrating the results of the automated analysis routine applied to the displacement traces. Average movement displacement and the frequency of production at each rate was calculated from the inferior-superior movement of the lower lip/jaw. Subjects repeated the syllables at a comfortable or preferred rate and as fast as possible for approximately six seconds. The peaks and valleys in the displacement signals are indicated by the vertical ticks above the traces and the summary measures were calculated as shown under each trace. From these results it can be seen that the lower lip/jaw movement for the more severe subject (S2) displays a smaller movement displacement compared to the less impaired subject (S1) although the preferred rate of repetition is approximately equivalent (2.9 vs. 2.8 Hz). At the fast rate the less severe subject (S1) is able to increase the frequency of production (2.8 to 5.4 Hz; 93% increase) with a concomitant reduction in the movement displacement (8.0 to 5.6 mm). In contrast, S2 is unable to increase the frequency of syllable repetitions to the same degree (2.9 to 3.5 Hz; 20% increase). In addition to measuring the movement displacement and frequency, similar measures can be made on the derived velocity time histories.

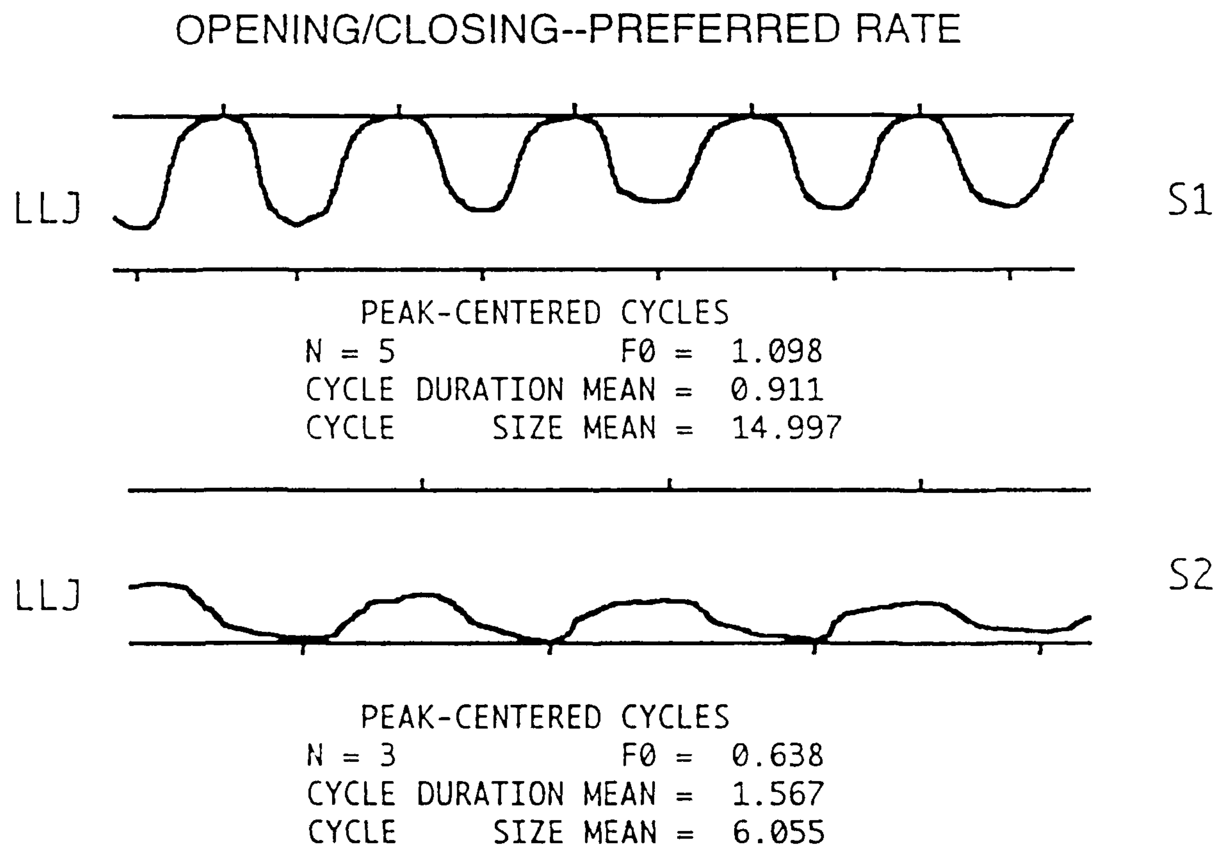

Figure 5. **Opening and closing lower lip/jaw movements in the inferior-superior direction for S1 and S2. Peak centered information is displayed under each trace.**

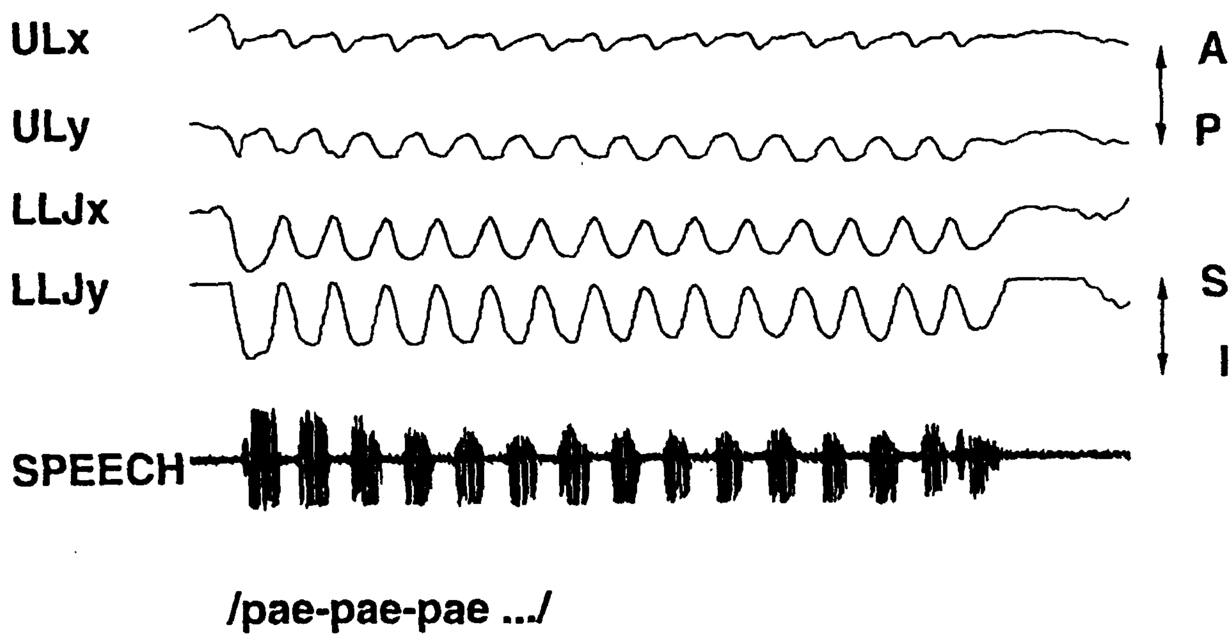

Figure 6. **Upper lip and lower lip/jaw movement in the anterior-posterior (x) and inferior-superior (y) directions for repetition of the syllable /pae/. As shown, UL and LLJ movement for the opening and closing involve movement in both x and y directions.**

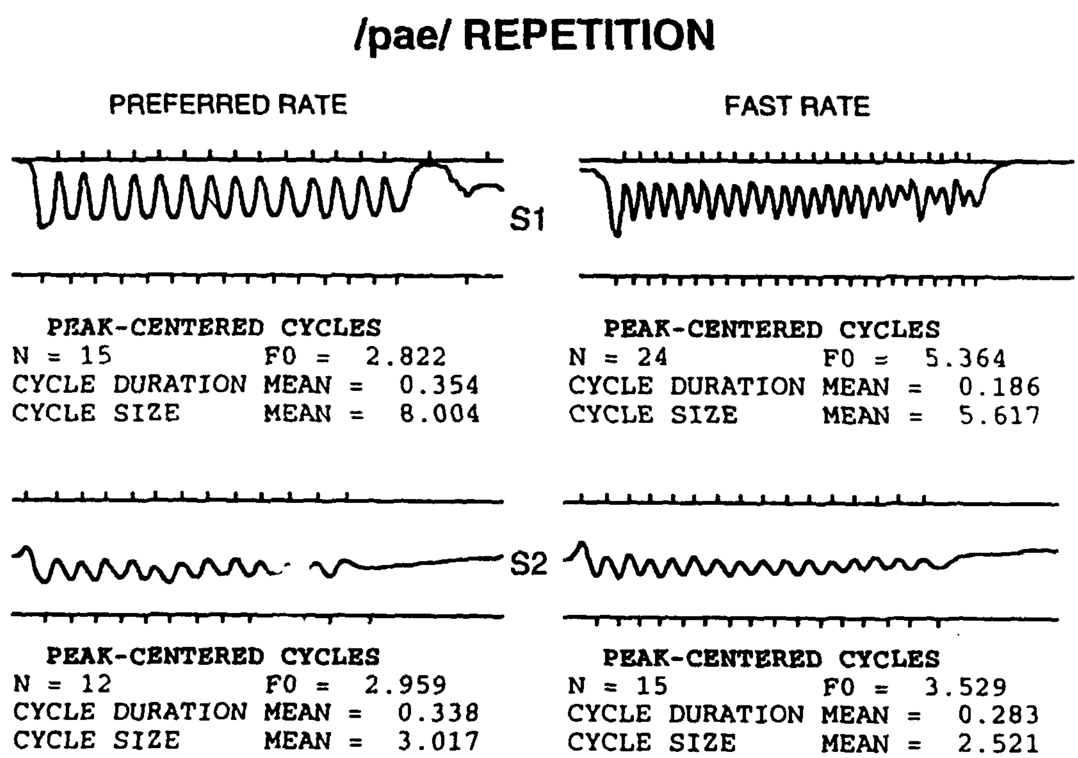

Figure 7. **Output from an automated analysis routine that picks peaks and valleys in the displacement signal (indicated by the vertical ticks above and below the respective signals) and calculates the number of peaks (valleys), the frequency of production (F0 in Hz), mean cycle duration (period), and the mean cycle size (mm). Shown are data from two subjects with different degrees of speech motor impairment secondary to PD. Subjects repeated the syllables at a preferred rate and as fast as possible.**

A comparison of speech and nonspeech movement tasks is presented in the next two figures (Figures 8 and 9). These data were obtained from the same two subjects presented in the previous figure. In each case the subject's task was to purse and retract the lips and to repeat the vowel sequence "uu"-"ee," at comfortable and fast rates. Because these movements are predominantly produced with anterior-posterior movements of the lips, only the anterior-posterior movements were measured. For S1 (Figure 8), both lips appear to be moving together (in phase) for all tasks. The consistency of the timing relations can be easily calculated using cross correlation. The nonspeech task (purse-retract) is not constrained by phonetic requirements and allows a more detailed evaluation of orofacial mobility. For this subject the nonspeech task is accomplished by equivalent contributions of the upper and lower lips. In contrast, "uu/ee" repetitions predominantly involve lower lip action. The frequency of both the speech and nonspeech tasks increase in the fast rate condition, although the nonspeech tasks demonstrates a greater degree of change. Results from the more severely involved subject (S2) are presented in Figure 9. For this subject, the rate changes are much less noticeable with the nonspeech task demonstrating a greater degree of impairment than was noted in the speech task. In addition, the nonspeech task was apparently difficult for S2 who demonstrates slow and labored protrusion and retraction of the lips. There is also some indication of a dyscoordination of the upper and lower lip movements at the faster rate during the speech task.

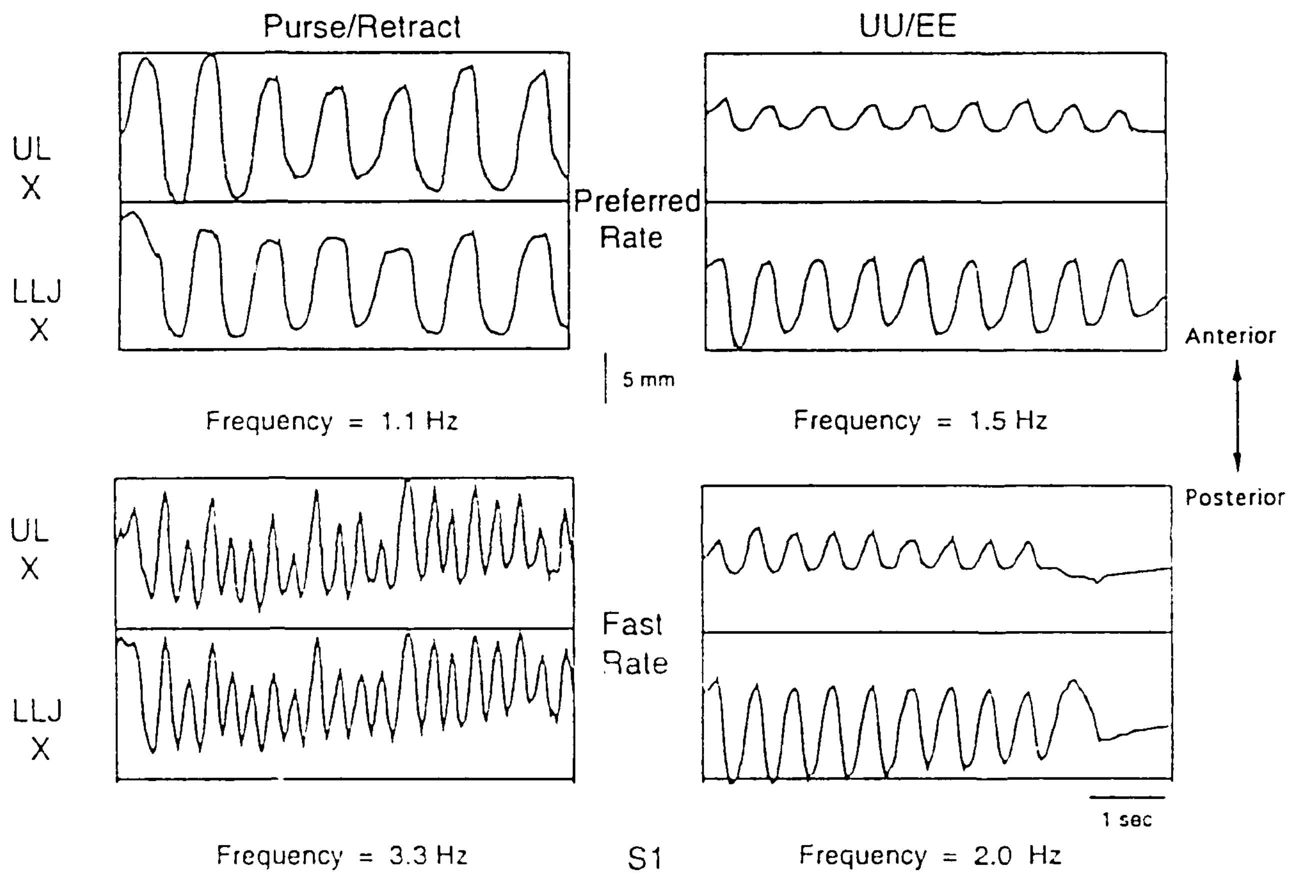

Figure 8. **Two different repetitive tasks involving predominantly anterior-posterior (x) motion of the UL and LLJ for S1. Shown are the position time histories for alternating and continuous pursing and retracting and alternating vowel production "uu-ee" at preferred and maximally fast rates. Below each panel is the average frequency of production.**

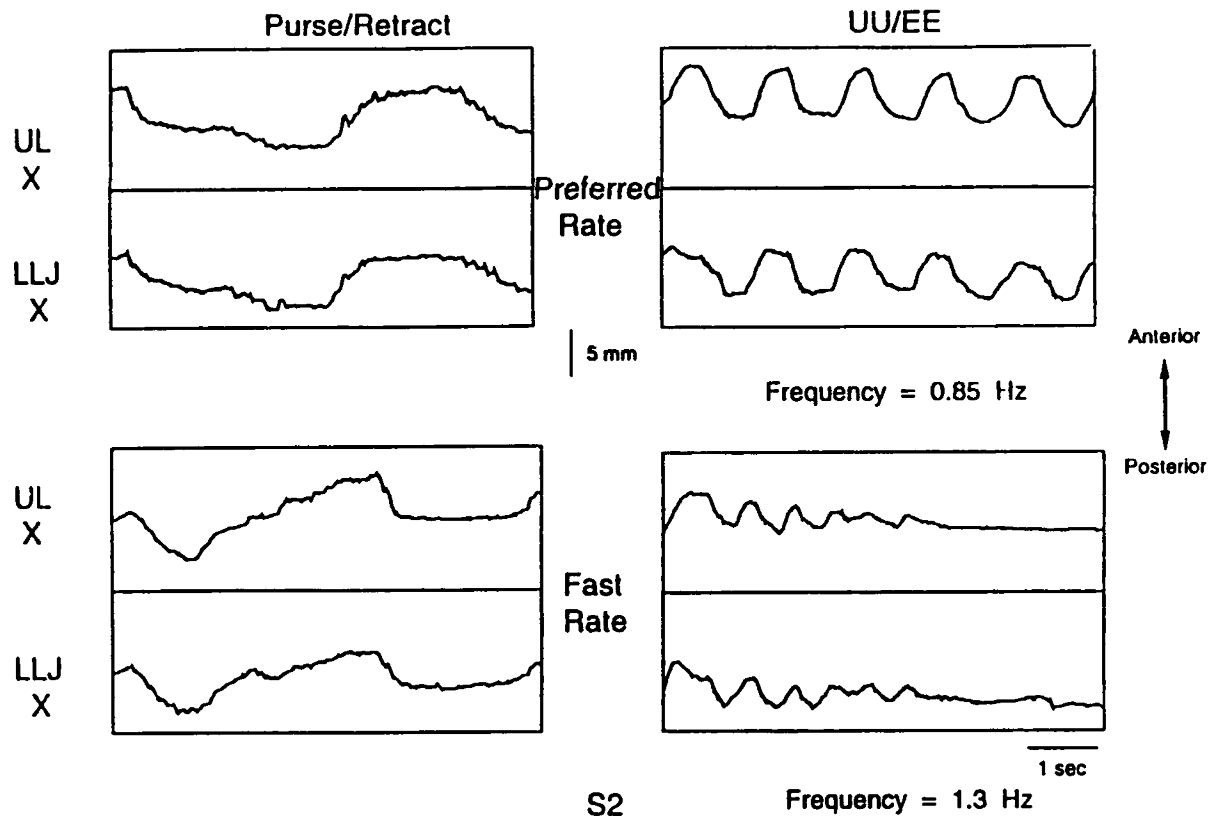

Figure 9. **The same repetitive tasks as in Figure 7 involving predominantly anterior-posterior (x) motion of the UL and LLJ for S2. Shown are the position time histories for alternating and continuous pursing and retracting and alternating and continuous vowel production "uu-ee" at preferred and maximally fast rates. Below each panel is the average frequency of production except for the purse/retract task because of the slowness of production.**

Other applications

There are additional applications in which movement transduction and analysis can used in the clinical evaluation of movement disorders. Instrumental tests can be used to provide information on the reaction time, speed, and visuomotor integrative abilities of the patient. In simple reaction time, the delay from the presentation of an auditory or visual stimulus to the onset of some response is measured. If the response involves movement to a target, such as closing the lips, the movement time can also be measured. Reaction time and movement time can be differentially affected in certain disorders such as Parkinsonism (Evarts, Teräväinnen, & Calne, 1981) and provide a means to objectively assess akinesia and bradykinesia, respectively during a nonspeech task. Tracking tests require the subject to follow a moving target with the output of a transducer attached to one of the articulators (see McClean, Beukelman, & Yorkston, 1987 for application to components of the speech motor system). Clinical applications usually involve scoring techniques which reflect the magnitude of the error between the target and the patients output. Such tests have been useful in evaluating ataxia or characterizing the impairment in producing smooth continuous motion of an effector. While not directly applicable to the perceptual deficits associated with speech motor disorders, these nonspeech results may prove useful in understanding the neurological condition, aspects of which may be masked by compensatory behavior of the patient. These novel techniques have been used in evaluating limb impairments associated with a variety of neurological disorders. The reader is referred to Potvin and Tourtellotte (1985) for an extensive compilation of measures and references.

RESEARCH NEEDS

In attempting to provide a quantitative basis for the evaluation of speech movements, two needs are obvious; the need for standardization and

normative data bases. All measures that have been described or implicated can be used to objectively monitor subject performance and evaluate disease progression or improvement due to therapeutic intervention. However, diagnostically, such measures have limited utility due to; 1) the lack of norms currently available, 2) solid correlational studies which attempt to relate kinematic characteristics with disease states or severity of involvement, and 3) technical standardization to allow valid intersubject comparisons. However, it may be the case that norms, while useful, may prove to be relatively uninformative or even misleading due to the range of variability in the normal population. This is not to suggest that normative data are not necessary. Rather, it may be more important to realize that speech movement data should not be evaluated in isolation without considering concomitant acoustic and perceptual characteristics of the disordered speech as well as overall motor and sensory performance levels. Only through a synthesis of observations can we hope to understand the communicative breakdown that is often interleaved with a more general sensorimotor deficit due to damage to the nervous system or modifications in nervous system operation.

CONCLUSIONS

Movement analysis is an objective and quantitative method of describing the behavior of the orofacial system during speech and nonspeech tasks. Evaluation of speech movement characteristics verify, refine, and extend, inferences and observations based on acoustic, aerodynamic, or auditory perceptual analyses. Both speech and nonspeech movements provide important information on the neuromotor functioning of the patient and facilitate assessment of disease states. Further, information related to the movement impairment can be easily assimilated by members of an interdisciplinary rehabilitation team. Quantitatively, movement transduction provides a reliable estimate of motor performance and can objectively monitor changes in performance associated with various forms of therapeutic intervention or changes in disease state. An improved understanding of movement deficits that underlie a specific motor disorder may lead to the development of novel treatment approaches that might not otherwise be considered.

REFERENCES

Abbs, J. H. & Gilbert, B. N. (1973). A strain gage transduction system for lip and jaw motion in two dimensions: Design criteria and calibration data. *Journal of Speech and Hearing Research, 19*, 19-35.

Abbs, J. H. & Watkin, K. L. (1976). Instrumentation for the study of speech physiology. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 41-75). New York: Academic Press.

Baken, R. J. (1987). *Clinical measurement of speech and voice*. Boston: Little, Brown, & Co.

Barlow, S. M., Cole, K. J., &, Abbs, J. H. (1983). A new head-mounted lip-jaw movement transduction system for the study of motor speech disorders. *Journal of Speech and Hearing Research, 26*, 283-288.

Chuang, C. K., & Wang, W. S.-Y. (1978). Use of optical distance sensing to track tongue motion. *Journal of Speech and Hearing Research, 21*, 482-496.

Evarts, E. V., Teräväinnen, H. T., & Calne, D. B. (1981). Reaction time in parkinson's disease. *Brain, 104*, 167-186.

Fletcher, S. G. (1982). Seeing speech in real time. *IEEE Spectrum*, April, 42-45.

Geddes, L. A., & Baker, L. E. (1968). *Principles of applied biomedical instrumentation*. New York: Wiley.

Gracco, V. L. (1990). Characteristics of speech as a motor control system. In G. Hammond (Ed.), *Cerebral control of speech and limb Movements* (pp. 1-28). Amsterdam: Elsevier.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience, 8*, 4628-4634.

Gracco, V. L., & Löfqvist, A. (1989). Speech movement coordination: Oral-laryngeal interactions. *Journal of the Acoustical Society of America, 86*, S114.

Gracco, V. L. & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Keller, E., & Ostry, D. (1983). Computer measurement of tongue dorsum movement with pulsed echo ultrasound. *Journal of the Acoustical Society of America, 73*, 1309-1315.

Kuehn, D. P., Reich, A. R., & Jordan, J. E. (1980). A cineradiographic study of chin marker positioning: Implications for the strain gauge transduction of jaw movement. *Journal of the Acoustical Society of America, 67*, 1825-1827.

McClean, M. D., Beukelman, D. R., & Yorkston, K. M. (1987). Speech-muscle visuomotor tracking in dysarthric and nonimpaired speakers. *Journal of Speech and Hearing Research, 30*, 276-282.

McCutcheon, M. J., Lakshminarayanan, A. V., & Fletcher, S. G. (1990). A comparison of glossometric measurements of tongue position with magnetic resonance images of the vocal tract. *Journal of the Acoustical Society of America, 87*, S122.

Michler, L., Bakke, M., & Møller, E. (1987). Graphic assessment of natural mandibular movements. *Journal of Craniofacial Disorders: Facial & Oral Pain, 1*, 97-114.

Müller, E. M., & Abbs, J. H. (1979). Strain gauge transduction of lip and jaw motion in the midsagittal plane: Refinement of a prototype system. *Journal of the Acoustical Society of America, 65*, 481-486.

Munhall, K. G., & Ostry, D. J. (1985). Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. C. Scherer (Eds.), *Vocal fold physiology: Biomechanics, acoustics and phonatory control* (pp. 145-162). Denver: Denver Center for the Performing Arts.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: [Human Perception and Performance], 11*, 457-474.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics, 46*, 135-147.

Ostry, D. J., Cooke, J. D., & Munhall, K. G. (1987). Velocity curves of human arm and speech movements. *Experimental Brain Research, 68*, 37-46.

Potvin, A. R., & Tourtellotte, W. W. (1985). *Quantitative Examination of neurologic functions. Volume I: Scientific basis and design of instrumented tests*. Boca Raton: CRC Press.

Sonies, B. C. (1982). Oral imaging systems: A review and clinical applications. *Journal of the National Student Speech Language Hearing Association*, *10*, 30-43.

Sonies, B. C. , Shawker, T. H., Hall, T. E., Gerber, L. H., & Leighton, S. B. (1981). Ultrasonic visualization of tongue motion during speech. *Journal of the Acoustical Society of America*, *70*, 683-686.

Stone, M., & Shawker, T. (1986). An ultrasound examination of tongue movement during swallowing. *Dysphagia*, *1*, 78-83.

Stone, M., Morish, K., Sonies, B., & Shawker, T. (1987). Tongue curvature: A model of shape during vowel production. *Folia Phoniatrica*, *39*, 302-315.

Stone, M., Shawker, T., Talbot, T., & Rich, A. (1988). Cross-sectional tongue shape during vowel production. *Journal of the Acoustical Society of America*, *83*, 1586-1596.

Sussman, H. M., & Smith, K. U. (1970a). Transducer for measuring mandibular movements. *Journal of the Acoustical Society of America*, *48*, 857-858.

Sussman, H. M., & Smith, K. U. (1970b). Transducer for measuring lip movements during speech. *Journal of the Acoustical Society of America*, *48*, 858-860.

Tuller, B., Shao, S, & Kelso, J. A. S. (1990). An evaluation of an alternating magnetic field device for monitoring tongue movements. *Journal of the Acoustical Society of America*, *88*, 674-679.

Winter, D. A. (1979). *Biomechanics of human movement*. New York: Wiley.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 59-72

Reiterant Speech as a Test of Nonnative Speakers' Mastery of the Timing of French*

Andrea Levitt†

The reiterant speech of ten native speakers of French was analyzed to develop baseline measures for syllable and consonant/vowel timing for a series of two-, three-, four-, and five-syllable French words spoken in isolation. Ten native speakers of English, who learned French as a second language, produced reiterant versions of both the French words and a comparable set of English words. The native speakers of English were divided into two groups on the basis of their second language experience. The first group consisted of four university-level teachers, who were relatively experienced learners of French, and the second group of six less experienced learners of French. The French reiterant imitations of the two groups of native speakers of English were compared to the native French speakers' productions. The timing patterns of the experienced group of non-native speakers did not differ significantly from those of the native French speakers, whereas there was a significant difference between these two groups and the group of six less experienced second-language learners. Deviations from the French baseline measures produced by the less experienced group are discussed in terms of the influence of the timing patterns of English and the literature on a sensitive period for second language acquisition.

INTRODUCTION

Although considerable research shows that native language phonetic habits influence second language productions, even for experienced second-language speakers (see Flege, 1986, for an extensive review), little work has been done on the influence of first language timing patterns on second language rhythmic patterns. One such study (Wenk, 1985) found an influence of native French rhythmic patterns on the timing of English as a second language. However, the effect of English timing patterns on the acquisition of French has not been directly tested.

The use of reiterant speech to test for such influence presents several advantages. In reiterant speech studies, subjects are asked to substitute a single syllable, often /ma/, for each of the original syllables in a word or sentence. Acoustic and perceptual analyses of reiterant speech have shown that it preserves the prosodic characteristics of the original utterance (Larkey, 1983; Liberman & Streeter, 1978; Nakatani, O'Connor, & Aston, 1981; Oller, 1973). Furthermore, because measurements of segment and syllable durations are easy with reiterant speech and are generally unconfounded by segmental variation, many studies have used such duration measurements in English for analyzing rhythm (e.g., Nakatani et al., 1981), for studying the perceptual effects of timing variations (Larkey, 1983; Nakatani & Schaffer, 1978), and especially for determining how durations vary as a function of utterance position and stress (e.g.,

This research was supported by NIH grant DC-00403 and NICHD Grant HD01994 and BRS Grant 05596 to Haskins Laboratories. I am grateful to Wellesley College for providing additional funding and to Leslie LeBarron and Katy Isaacs for conducting preliminary statistics on some of the early durational measurements. Mindy Levitt provided valuable assistance with data entry, and Tony Hatoun supplied special statistical programs. The French subjects were run while the author was supported by the North Atlantic Treaty Organization under a grant awarded in 1983 and held at the Centre National de la Recherche Scientifique in Paris. I thank Jacques Mehler for making available the facilities of his laboratory. Catherine Best, Jane Baier, Jim Flege, Bob Frye, Alice Healy, Patti Price, Bruno Repp, Michael Studdert-Kennedy, Doug Whalen, and an anonymous reviewer provided helpful comments on previous versions of this manuscript.

Oller, 1973). Reiterant speech duration measurements have also been made on Swedish (e.g., Lindblom & Rapp, 1973), and comparisons of the rhythmic features of a group of languages have been made on the basis of reiterant speech (Hoequist, 1983; Vatikiotis-Bateson, 1986). However, very little work has been done with reiterant speech on the rhythmic features of French, aside from that done by Vatikiotis-Bateson (1986), where reiterant speech was used to determine universal and language-specific effects on articulator timing in native speakers from a group of languages. The use of reiterant speech as a means of testing a non-native speaker's mastery of the timing patterns of a foreign language has not been previously attempted. In learning a second language, speakers need to learn new timing patterns for individual segments, often as a function of context (Mack, 1982), as well as new rhythmic patterns. Reiterant speech is particularly well suited to testing the acquisition of new rhythmic patterns independently from the effects of timing for non-native segments.

The speech rhythm of French and that of English are quite distinct. French has been traditionally classified as a "syllable-timed" language (e.g., Pike, 1945), with syllables essentially equal in length. This characterization of French rhythm has been criticized (e.g., Dauer, 1983; Fletcher, 1991; Wenk & Wioland, 1982) for failing to recognize the important final-syllable lengthening that is characteristic of French rhythmic groups, which may be either the individual "sense groups" of a French sentence or individual French words spoken in isolation. Thus, nonfinal syllables within unemphatic French rhythmic groups are, except for effects of phonetic variation, essentially equal in length, whereas final syllables show considerable lengthening. English, on the other hand, has been traditionally classified as a "stress-timed" language (e.g., Pike, 1945). Because of variable word stress, any English sentence presents a series of stressed syllables which alternate with unstressed syllables. A stress-timed language is supposed to maintain equal intervals between stressed syllables. Thus, if an interval between two stressed syllables contains more unstressed syllables than another, those unstressed syllables should show relatively greater compression. English also exhibits characteristic patterns of final-syllable lengthening, including word-final, phrase-final, and utterance-final lengthening (Oller, 1973).

Although the characterization of English as a "stress-timed" language has also been criticized (e.g., Dauer, 1983; Wenk & Wioland, 1982), its rhythmic pattern is nonetheless quite different from that of French, especially in two salient respects. First, in English, nonfinal syllables will vary in length as a function of stress, whereas in unemphatic French, nonfinal syllables within a rhythmic group are essentially equal in length. Second, although both languages exhibit final-syllable lengthening, the magnitude of the final-syllable lengthening effect and its location both vary. Thus, the magnitude of utterance-final lengthening is greater in French than in English (Delattre, 1966). In addition, in English, utterance-final lengthening appears to be greater than phrase- or word-final lengthening (e.g., Oller, 1973). A similar difference in the magnitude of final-syllable lengthening has been observed for utterance-final compared to phrase-final lengthening in French (Benguerel, 1971; Fletcher, 1991; but cf. Allen, 1973), but not for words. French words exhibit final lengthening only at the ends of rhythm groups or when uttered in isolation.

Which of these rhythmic differences are second-language learners of French likely to master first? On the one hand, since both languages exhibit final-syllable lengthening, English-speaking learners of French might find it easier to adjust the magnitude of such lengthening as they acquire the rhythm of French. On the other hand, Flege (e.g., Flege, 1981; Flege, 1987; Flege & Hillenbrand, 1984) has proposed that second-language learners are more likely to master the totally new phonetic features of a second language than those that can be assimilated to their native repertoire. In that case, English-speaking learners of French might find it easier to acquire the relatively equal timing of nonfinal syllables in French, which is not found in English.

In order to conduct a test of the acquisition of French rhythmic patterns by native speakers of English, it is first necessary to establish baseline measures for timing patterns in French using the reiterant productions of native speakers of French. Not all speakers are equally good at producing reiterant speech that preserves the timing of the original utterance (Larkey, 1983). Thus, it is important that the baseline measures be based on the fluent productions of the best reiterant speakers. Once these measures have been established, they can be compared to published findings about the durations of consonants, vowels, and syllables in French.

Experiment I reports the results of an experiment designed to produce such data.

We may then ask how well non-native speakers of French match the timing patterns of the native French productions. In Experiment II, reiterant versions of both French and English words made by native speakers of English were analyzed in order to establish a similar set of baseline measures for reiterant English, to determine how well the non-native speakers of French differing in degree of experience with the language match the timing of the productions of the French speakers, and to see whether any deviations from the French baseline measure stem from the influence of English timing patterns.

I. EXPERIMENT 1

A. Subjects. Ten subjects, five male and five female, participated in the study. All were native speakers of French from the Paris region. All of the subjects have advanced graduate degrees. Although the majority of their daily verbal exchanges took place in French, all the subjects had some experience with other languages, as is typical of highly educated Europeans.

B. Test materials. The materials for the experiment consisted of a set of 30 French words, 6 two-syllable, 12 three-syllable and 6 each of four- and five-syllable words. (See the Appendix for a complete list of the stimuli.) As stress in French is on final syllables, and all of the words were produced in isolation, all of the two-, three-, four-, and five-syllable words in French were stressed on their final syllable. Each word was typed on the center of a 3 × 5 card. The cards were presented in the same random order to all subjects.

C. Procedure. Recordings were made in a soundproof booth using a Sony tape recorder (model TC-510-Z) and a Sennheiser microphone (model MD 441-V). The subjects read the word typed on the card out loud and then reproduced what they had just said by substituting the syllable /ma/ for every syllable of the original, while preserving both its timing and the melodic contour. They were asked to be careful to use the syllable /ma/ in all cases and to repeat a stimulus item and its reiterant version, if they felt they had made an error.

D. Equipment and measurement methods. The 30 French words and their reiterant versions were low-pass filtered at 4.9 kHz, digitized at 10 kHz, and stored on disk, using Haskins Laboratories' Vax 11-780 computer. All durational measurements were made by the author on the reiterant speech using large-scale waveform displays, with a resolution of 0.1 ms. Differences in amplitude between the consonant and the vowel, as well as differences in the appearance of the waveforms associated with /m/ (the nasal murmur) and /a/, made segmentation relatively easy. This was particularly true for reiterant productions by French speakers. It was very easy in almost all cases to segment the /m/ and the /a/ because French /m/ and /a/ are kept quite distinct, whereas English oral vowels in a nasal environment often show some nasalization (Clumeck, 1975). When there was a question about the location of a particular boundary, it was resolved through listening to the segments in question. The most common segmentation difficulty arose in determining the location of the end of the word. A consistently conservative criterion was applied, such that the termination of periodicity was used to mark the end point. This excluded breathy releases, but seemed best for consistent comparisons across speakers.

In order to test the reliability of the duration measurements, a random sample of 12 French reiterant utterances containing 82 separate measurements were measured a second time by the author. Absolute duration measurement differences were within 4 ms of the original on the average overall and within 9 ms on the average on the 12 final vowel measurements.

Not all individuals are equally adept at producing reiterant speech that faithfully mimics the prosodic characteristics of the original utterances. To construct accurate timing models, we must require that the reiterant utterances chosen for analysis come from subjects who have demonstrated that they are capable of neutralizing inherent segmental length differences. That is, the subject must produce reiterant syllables of the same length, all other things being equal, for both original syllables that are inherently long and for ones that are inherently short. Reiterant speech studies typically use specially constructed sentences that are rhythmically matched, based on their stress patterns, although one sentence of each pair contains words with inherently long syllables and one sentence contains words with inherently short syllables. Thus, the sentences in each pair are rhythmically the same, with the same number of syllables and the same locations for stressed syllables, but the individual syllables vary in length. Subjects should produce essentially identical reiterant productions for both sentences

in a set, if, in fact, they are neutralizing intrinsic differences in the durations of individual segments.

In the present study, each of the two-, three-, four- and five-syllable word-length types had syllables composed of segments of inherently different lengths. Thus, instead of using a sentence-length test, measures of subjects' duration measurement variability in producing word types were used as an indication of their ability to neutralize inherent segmental length differences. Each reduplicative version of a particular word of a given length was considered a token of that word-length type. The standard deviations for comparable measurements, e.g., first syllable length, were calculated across tokens for each subject for each word-length type and averaged. Separate values were calculated for each of the four word-length types because it is generally more difficult to produce good reiterant versions for longer utterances. Finally, an overall mean (measure A) and a standard deviation (measure B) of each subject's mean standard deviations for the four French word-length types were calculated. The overall group mean was 25 ms for measure A and 20 ms for measure B. Subjects were rank ordered on both measures, and three subjects, one female and two males, showed means and standard deviations that were consistently longer than the other subjects (35 ms for measure A and 33 ms for measure B for the group of three). They had also produced more errors between them (16) than the other seven subjects combined. Their data were excluded from the construction of the French baseline measures for timing. For the remaining seven subjects, the mean for measure A was 19 ms with a mean 14 ms for measure B.

E. Results. There were only seven errors made across the seven subjects (3%), most of which involved the addition or deletion of a syllable, usually on words of four or five syllables. All errors were excluded from the construction of the baseline measures for timing. There were also two instances of missing data (1%).

Figure 1 shows the mean durational measurements for the syllables of the reiterant versions of each of the four word-types in terms of the mean durational measurements of the consonants (/m/) and vowels (/a/) of each syllable. The mean duration of /m/ in nonfinal syllables was 83 ms, of /m/ in final syllables was 103 ms, of /a/ in nonfinal syllables was 93 ms and of final /a/ was 171 ms. Nonfinal syllables averaged 175 ms in length, whereas final syllables measured 274 ms on the average, an increase of almost 100 ms or a final/nonfinal ratio of 1.6.[1]

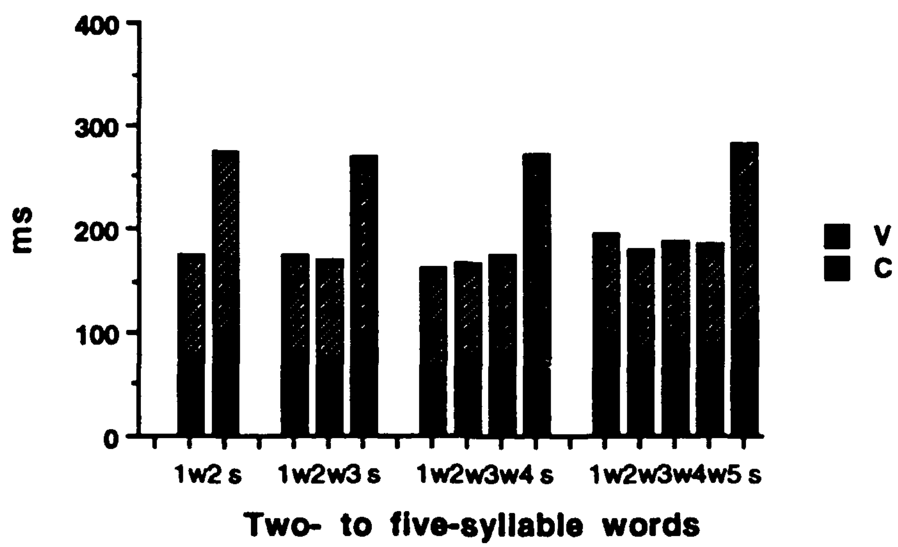

Figure 1. **Consonant and vowel durations, as a function of word length, syllable position, and stress, for reiterant productions of French words spoken in isolation by native speakers of French. (Numbers indicate syllable position, S indicates stressed syllables, and W indicates unstressed syllables).**

This final-syllable lengthening was found to be significant in the results of a two-way analysis of variance comparing the subjects' mean nonfinal and final syllable lengths for the four word-length types [$F(1,6)=130.19$, $p < 0.0000$]. There were no word-length type and no word-length type by syllable position interactions. Analyses comparing subjects' mean nonfinal syllable lengths for each of the four word-length types were also not significant.[2] A separate two-way analysis of variance to explore segment length in final and nonfinal syllables again showed a highly significant effect of syllable position [$F(1,6)=105.8$, $p < 0.0000$]. There was also a significant effect of segment type [$F(1,6)=46.01$, $p < .0005$], and a syllable by segment type interaction [$F(1,6)=60.26$, $p < 0.0002$]. Post hoc tests (Newman-Keuls) revealed that final /a/ was significantly different from nonfinal /a/ and from final and nonfinal /m/ and that final /m/ was significantly different from nonfinal /m/, all at the $p < 0.05$ level or better. Nonfinal /m/ and /a/ were not significantly different from one another.

Table 1 shows the mean length of each of the word types and the ratio of the mean length of the consonant to that of the vowel in each syllable. The overall mean C/V ratio was .9 for nonfinal syllables and the C/V ratio was .6 for final syllables. In addition, Table 1 presents the ratios of the mean syllable length to the word as a whole.

Table 1. *Mean word lengths (in ms) and C/V and CV/length ratios in reiterant speech productions of French words by native speakers of French.*

	Word Length in Syllables			
	Two	Three	Four	Five
Mean Word Length	448.2	611.6	776.2	1027.5
Ratios				
C1/V1	.9	.9	.8	.9
C2/V2	.6	.8	.9	1.0
C3/V3		.6	.9	1.0
C4/V4			.7	.9
C5/V5				.7
Ratios				
CV1/L	.4	.3	.2	.2
CV2/L	.6	.3	.2	.2
CV3/L		.4	.2	.2
CV4/L			.4	.2
CV5/L				.3

F. Discussion

The results of this experiment showed fairly good agreement with the published data on French, especially with respect to French syllable duration ratios. The segment measurements will be considered first and then the syllable measurements.

The duration measurements for French nonfinal /m/ and /a/ and for final /m/ tended to be roughly 20 ms longer than the durations found for the same segments by other researchers (Di Cristo, 1980; O'Shaughnessy, 1984; Smith, 1977). This discrepancy is most likely due to the fact that the subjects in the present experiment spoke at a slower rate in producing reiterant speech than the subjects in the other studies, who read French texts. The measurement for utterance-final /a/ was roughly 10 ms longer than that of O'Shaughnessy (1984). The smaller discrepancy in final position is probably due to the conservative segmentation criterion adopted in the present study. Thus, given the segment values of the present study, the nasal consonant /m/ accounted for 47% of the duration of nonfinal syllables, whereas for final syllables, it accounted for 38%.

In general, nonfinal syllables were remarkably close in duration (see Figure 1). The present data did not show an initial syllable shortening as compared to medial syllables, which disagrees with Crompton's (1980) finding of decreased length for initial syllables. In fact, another researcher (Vaissière, 1983) has found growing evidence in French of a tendency to stress word initial syllables, and presumably to lengthen them. Indeed, one of the subjects showed a regular lengthening of initial syllables. Crompton (1980) also found evidence for prenuclear lengthening, or lengthening of a syllable just prior to a nuclear stress. An analogous penultimate syllable lengthening has been described by Smith (1977) as characteristic of Parisian French (although only one of Crompton's four subjects was from Paris, while the other three came from Brittany). The present pooled data show no overall effect of penultimate syllable lengthening, although data from two of the speakers do show such an effect.

The ratio of final syllable to non-final syllable length in the present data was 1.6, which agrees exactly with Parmenter and Blanc's measure of 1.6 (1933), with Benguerel's (1971) measure of 1.6, and with Allen's (1983) finding of an overall ratio of 1.6 when he compared the median lengths of final to penultimate vowels in French children's productions of French words. It does not match Delattre's (1966) measure of 1.8, perhaps because

of differences in the criteria used for measuring final syllable lengths.

In summary, our French timing data based on reiterant speech productions of French words spoken in isolation showed generally consistent syllable durations for nonfinal syllables and a ratio of final/nonfinal syllables of 1.6. Individual subjects showed some slight lengthening of initial or penultimate syllables, but no consistent evidence for any shortening effects. Insofar as intrasyllabic timing is concerned, in nonfinal syllables, the nasal accounted for 47% of the duration, and in final syllables, it accounted for 38%. How well then do non-native speakers of French match these characteristic duration patterns when they produce reiterant speech versions of French words?

II. EXPERIMENT 2

A. Subjects. Ten subjects, five male and five female, participated in the study. All of the subjects except for one have advanced graduate degrees. All are native speakers of English, currently living in the Boston area, who have studied standard French. Four of the subjects (two men and two women, including the author) teach French at the university level. One subject learned French from his French wife, whom he met after graduate school. The other subjects all had some formal training in French; seven subjects began the study of French in high school and the remaining two in junior high school. The four teachers of French and the other subjects, with the exception of the subject who learned French at home, averaged over two years of high school French. The four French teachers, however, studied French for four years in college, as compared to an average of slightly over 1 1/2 years in college for the others. The four French teachers also completed postgraduate training in French and had traveled more extensively in French-speaking countries than had the other subjects.

B. Test materials. The same French deck of 3 × 5 cards used in the previous experiment was used in this second study. An additional deck consisting of the English cognates of the French words was also used. The 30 English words consisted of two, three, four or five syllables. There were ten possible stress patterns represented. For words of two syllables, both initial and final primary stress patterns occurred (*sacred* and *degree*.) For words of three syllables, initial, medial and final primary stress patterns occurred (*compliment*, *instructive*, and *engineer*). For words of four syllables, three of the four possible primary stress patterns occurred (*commentary*, *economy*, and *exposition*). For words of five syllables, two possible patterns occurred (*electricity* and *communication*). There were three different words representing each of the syllable and stress types.[3] Although in general most of the cognates had the same number of syllables in the two languages, there were three items for which the syllable count differed. (See the Appendix for a complete list of the stimuli used).

C. Procedure. Subjects first filled out a short questionnaire about their years of experience with French and were then recorded in a quiet room, onto a Teac tape recorder (model X-7MKII) using a Realistic dynamic microphone (model 33-984A). The rest of the procedure was the same as in the previous experiment, except that subjects read and produced reiterant versions the words of the English deck first.

D. Equipment and measurement methods. All 30 French and 30 English words and their reiterant versions were low-pass filtered at 4.9 kHz, digitized at 10 kHz, and stored on disk on Haskins Laboratories' Vax 11/780. The same criteria used in the previous experiment were used here to determine the consonant and vowel boundaries and the end of the reiterant speech utterance.

A random sample of fourteen reiterant productions of English words containing 102 separate measurements were measured a second time. The absolute duration measurements were within 4 ms of the original measures on the average overall, and within 9 ms on the average for the fourteen final vowel measurements.

The errors from both sets of reiterant productions will be discussed first. The data from Experiment 2 will then be presented as a set of baseline measures for consonant, vowel, and syllable timing for English words of various lengths and stress patterns based on the productions of the most consistent reiterant speakers. Third, the English speakers' reiterant versions of the French words will be examined for patterns of intra- and intersyllabic timing. Finally, the durations of the productions of the French native speakers will be statistically compared to those of the non-native speakers, broken into two groups, the relatively experienced teachers of French and the other, less experienced group of French learners.

As with the French subjects, measures of the American subjects' duration measurement variability in producing word types were used as an indication of their ability to neutralize inherent segmental length differences. Each reduplicative version of a particular word of a given length and stress pattern was considered a token of that

word-length/stress-pattern type. The standard deviations for comparable measurements, e.g., first syllable length, were calculated across tokens for each subject for each of the ten word-length/stress-pattern types and averaged. Separate values were calculated for each of the ten word-length/stress-pattern types because it is generally more difficult to produce good reiterant productions for longer utterances and because variable word stress in English affects the duration of syllables in comparable positions. Finally, an overall mean (measure A) and a standard deviation (measure B) of each subject's mean standard deviations for the ten word-length/stress-pattern types were calculated. For the English words, the group mean on measure A was 18 ms with a group mean on measure B of 17 ms. When the subjects were rank ordered on these two measures, two subjects, one male and one female, showed the highest scores on both measures (for measure A, their mean was 26 ms, with a mean of 24 ms for measure B). The remaining eight subjects showed a group mean of 17 ms on measure A and 15 ms on measure B. In constructing the baseline measures for timing for the English words, only the data from the eight most consistent subjects were included.

E. Results

The American subjects made relatively few errors in their reiterant versions of the English words. The twelve errors across the eight most consistent subjects gave an error rate of 5%, with most errors due to a subject's producing an incorrect number of syllables for one of the longer words or to a subject's clearly stressing the wrong syllable in the reiterant production. There were only two missing tokens (.8%). The American subjects made many more errors in their reiterant versions of the French words. There were twenty-nine such errors (12%) across the eight subjects. Twenty-four of those errors (83% of the total), were words ending in "ion" or containing the vowel sequence "ié" as in "société," which the French count as a single syllable, but which many of the Americans counted as two. There was only one missing token (.4%).

Figure 2 presents the averaged durational measurements of the eight American speakers for each of the ten word types as a function of the consonants (/m/) and vowels (/a/). For initial stressed syllables,[4] /m/ averaged 56 ms and /a/ 92 ms, for medial stressed syllables, /m/ averaged 79 ms and /a/ 108 ms, for final stressed syllables, /m/ averaged 82 ms and /a/ 255 ms. For unstressed syllables, /m/ averaged 45 ms and /a/ 70 ms in initial syllables, /m/ was 65 ms and /a/ was 76 ms in medial syllables, and /m/ was 79 ms and /a/ was 155 ms in final syllables. The mean duration of syllables bearing primary stress[5] were 160 ms in initial position, 190 ms medially, and 336 ms finally. Syllables with secondary stress averaged 137 ms initially and 168 medially. Syllables that were not stressed averaged 113 ms initially, 138 ms medially and 233 ms finally.

Table 2 shows the overall mean length for each word type, the consonant/vowel ratios for each syllable and the ratios of each of the individual syllables to the length of the word.

Figure 3 shows the mean durational measurements for the reiterant versions of the syllables of each of the four French word-length types, as produced by the native speakers of English, in terms of consonants (/m/) and vowels (/a/). The mean duration of /m/ in nonfinal syllables was 73 ms, of /m/ in final syllables was 95 ms, of /a/ in nonfinal syllables was 85 ms, and of /a/ in final syllables was 235 ms. Nonfinal syllables thus averaged 157 ms, whereas final syllables averaged 330 ms. The difference in syllable length averaged over 170 ms and produced a final/nonfinal ratio of 2.1.

The results of a two-way analysis of variance comparing the subjects' mean nonfinal and final syllable lengths for the four word-length types showed a highly significant effect of syllable position [$F(1,9)=182.22$, $p < 0.0000$], but no word-length type and no word-length type by syllable position interaction. Separate analyses comparing subjects' mean nonfinal syllable lengths for each of the four word-length types were also not significant.[6]

Table 3 shows the mean length of each of the word-length types and the ratio of the mean length of the consonant to that of the vowel in each syllable. The overall mean C/V ratio was .9 for nonfinal syllables, which was comparable to that of the French subjects, but the overall mean C/V was .45 for final syllables, which was different from that of the French subjects.

In order to test how well the American subjects conformed to the French baseline measures for timing for nonfinal and final syllables in their reiterant productions of French words, their timing measures were subjected to an analysis of variance with one between group factor with three levels (native French versus teachers of French versus English speakers) and two within group factors (syllable position [nonfinal versus final] and segment duration [consonant versus vowel length]).

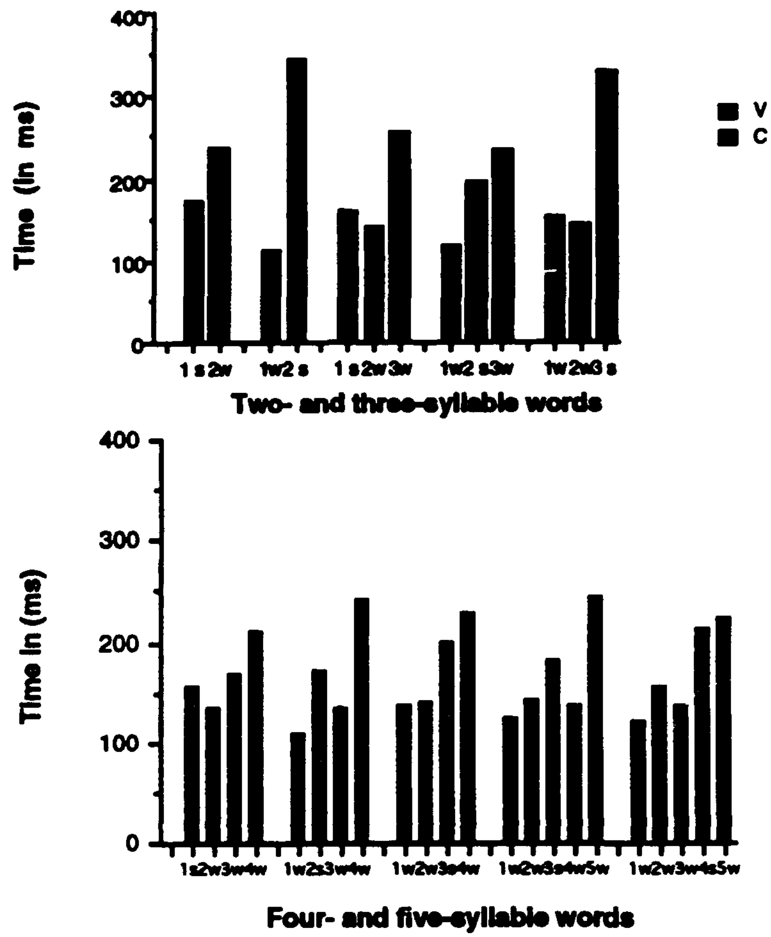

Figure 2. **Consonant and vowel durations, as a function of word length, syllable position, and stress, for reiterant imitations of English words spoken in isolation by native speakers of English. (Numbers indicate syllable position, S indicates stressed syllables, and W indicates unstressed syllables or those bearing secondary stress).**

Table 2. *Mean word lengths (in ms) and C/V and CV/Length ratios in reiterant speech productions of English words by native speakers of English.*

	Word Length in Syllables									
	Two		Three			Four			Five	
Stress Type	1	2	3	4	5	6	7	8	9	10
Mean Word Length	408.0	457.1	552.4	542.5	624.4	663.0	651.3	703.3	825.7	861.5
Ratios										
C1/V1	.5	.6	.6	.6	.7	.7	.6	.6	.6	.7
C2/V2	.5	.3	.9	.7	.9	.8	.8	.9	.8	.9
C3/V3			.5	.6	.4	.9	.8	.7	.8	.8
C4/V4						.8	.5	.5	.8	.7
C5/V5									.6	.5
Ratios										
CV1/L	.4	.2	.3	.2	.2	.2	.2	.2	.1	.1
CV2/L	.6	.8	.3	.4	.2	.2	.3	.2	.2	.2
CV3/L			.5	.4	.5	.3	.2	.3	.2	.2
CV4/L						.3	.4	.3	.2	.2
CV5/L									.3	.3

Tokens of types: 1=counter; 2=control; 3=compliment; 4=conclusion; 5=engineer; 6=commentary; 7=economy; 8=exposition; 9=elasticity; 10=communication.

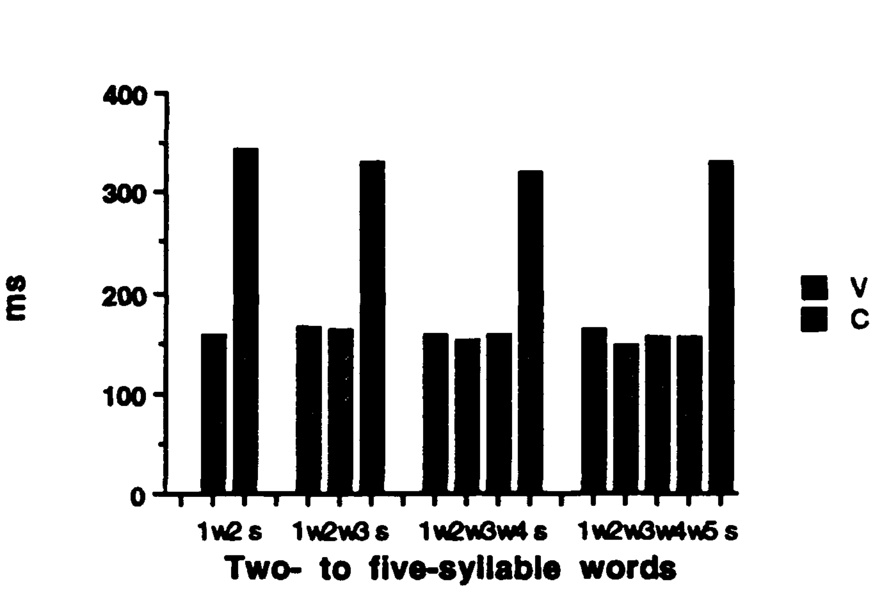

Figure 3. **Consonant and vowel durations, as a function of word length, syllable position, and stress, for reiterant French words spoken in isolation by non-native speakers. (Numbers indicate syllable position, S indicates stressed syllables, and W indicates unstressed syllables).**

Table 3. *Mean word lengths (in ms) and C/V and CV/Length ratios in reiterant speech productions of French words by native speakers of English.*

	Word Length in Syllables			
	Two	Three	Four	Five
Mean Word Length	500.5	656.1	786.1	943.6
Ratios				
C1/V1	.7	.8	1.0	1.0
C2/V2	.4	.9	1.0	.9
C3/V3		.4	.8	1.0
C4/V4			.5	.9
C5/V5				.5
Ratios				
CV1/L	.3	.2	.2	.2
CV2/L	.7	.3	.2	.2
CV3/L		.5	.2	.2
CV4/L			.4	.2
CV5/L				.2

Although there was no significant main effect of group, there was a significant effect of syllable position [$F(1,17)=417.87$, $p < 0.0000$] and of segment duration [$F(1,17)=121.42$, $p < 0.0000$], and both of these effects interacted significantly with the group factor [$F(2,17)=15.41$, $p < 0.0003$], in the case of syllable position, and [$F(2,17)=8.28$, $p < .0032$], in the case of consonant versus vowel length. There was also a significant two-way interaction of syllable position and segment duration [$F(1,17)=145.20$, $p < 0.0000$] that also interacted significantly with the group factor [$F(2,17)=10.88$, $p < 0.001$]. Figure 4 shows the pattern of results for the three groups.

An exploration of the group interactions with syllable position and consonant versus vowel revealed that the source of the interactions was the differences in final syllable length among the three groups, in particular due to differences in the vowel length, as can be seen in Figure 4. A separate analysis of variance conducted on final syllable vowel length was significant [$F(2,17)=7.65$, $p < .0044$]. Post hoc (Newman-Keuls) tests revealed that in terms of final vowel length, the productions of the native speakers of French and the French teachers did not differ from one another but the productions of both groups differed from those of the other native English speakers ($p<.05$).

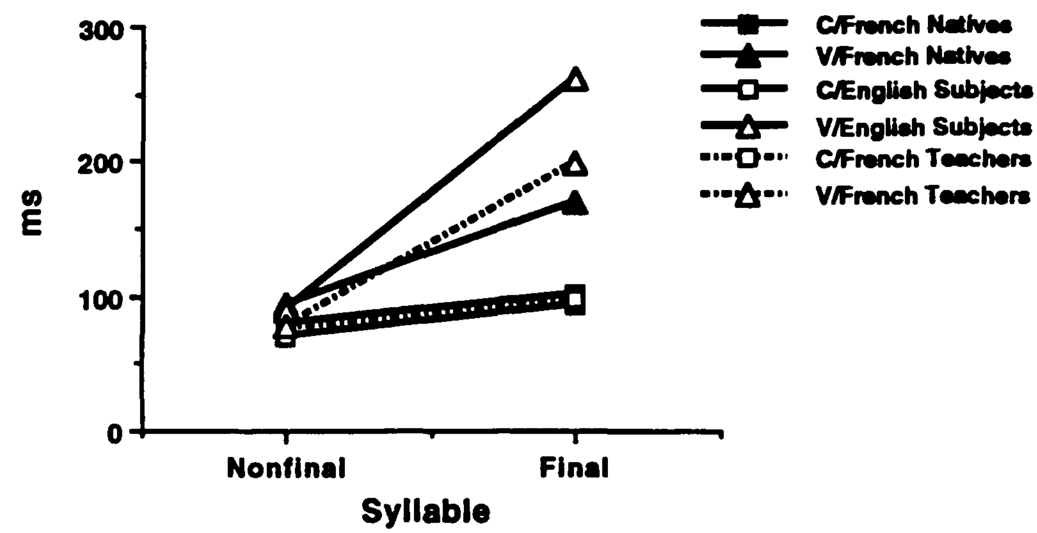

Figure 4. Mean consonant and vowel length for final and nonfinal syllables for French native speakers, experienced learners of French (French Teachers), and relatively inexperienced learners of French (English Subjects).

F. Discussion

The American subjects' productions of the English segment and syllable durations will first be discussed, followed by an examination of the ways in which their reiterant productions of the French words deviate from the French baseline measures. Finally the possible effects of English timing patterns on the French productions will be considered.

In the English reiterant speech, the nasal murmur accounted for 38% of the syllable in stressed initial syllables, 42% in stressed medial syllables and 24% in stressed final syllables. For unstressed syllables the percentages were 39% initially, 45% medially and 34% finally. These percentages clearly differ from those found in French in Experiment 1, which suggests that the intrasyllabic timing is not the same in the two languages.

There was also clearly an effect of utterance-final lengthening carried largely by the vowel in the English data. For stressed syllables, lengthening for final vowels was roughly 150 ms and for unstressed syllables it was roughly 75 ms. These durational lengthenings are comparable to those found by Oller (1973).[7]

Insofar as the syllable measurements are concerned, the present data showed clear effects both of stress and of utterance-final lengthening. There also appeared to be increments due to secondary stress, although Nakatani et al. found only marginal increases in length for such syllables and only for some speakers. The ratio of final/nonfinal syllables was 1.7, which is greater than the 1.5 found by Delattre (1966), but which may be due to the unusually short initial syllables found in this study. Indeed, if initial syllables are eliminated from consideration, the ratio becomes 1.6, which is closer to Delattre's measure. The ratio of accented to unaccented syllables was 1.43 in initial syllables, 1.38 in medial syllables and 1.44 in final syllables. These ratios, which do not include the somewhat problematic syllables that bear secondary stress, correspond fairly well to Hoequist's measure of 1.45, although they are lower than the measure given by Delattre (1966) of 1.7. Hoequist's (1983) suggestion that Delattre's higher ratio is due to the inclusion in the unstressed group of very short /ə/ syllables, which are generally not found in reiterant speech, seems quite reasonable.

As can be seen in Figure 4, for the reiterant versions of the French words, there was little difference in the consonant and vowel lengths in nonfinal syllables for the three groups. Thus, the percentage represented by the nasal in nonfinal syllables was 47% for the native speakers of French, 49% for the American teachers of French, and 44% for the less experienced French speakers. There was also little difference in the mean length of /m/ in final syllables for the three groups of subjects. The striking difference in the reiterant productions of the three groups occurs in the length of utterance-final /a/ which was 171 ms for the French natives, 199 ms for the French teachers, and 260 ms for the less experienced

group. Thus, the nasal consonant accounts for 38% of the final syllable for French natives, 33% for French teachers, and only 26% for the less experienced group. Intrasyllabic timing appears to be more native-like in nonfinal than in final syllables. The ratio of final to nonfinal syllables was 1.6 for the French natives, 1.9 for the French teachers, and 2.2 for the others. Although the reiterant productions of the American teachers of French were not significantly different from those of the French natives, in almost all cases, the teachers' productions, while close to those of the French natives, fall between that group and the other group of native speakers of English.

Surprisingly, the Americans had a durational pattern in their reiterant versions of English words that turned out to be very close to the French timing pattern. Thus, the average duration of the first syllable in two syllable words with stress on the first syllable (see Figure 2) was 173 ms while the final syllable was 236 ms on the average, which is comparable to the French natives' 176 ms average length for nonfinal syllables and 274 ms average length for final syllables. Yet many of the Americans who were less experienced in French seemed to match the durational pattern of the final syllable of French words uttered in isolation (353 ms) by patterning it after the duration of their own stressed syllables in final position (336 ms) whereas the teachers of French achieved a closer match to the French baseline measure (296 ms).

Insofar as the nonfinal syllables are concerned, all the Americans showed that they can generally produce syllables of quite equal length (see Figure 3), and there was no indication in their reiterant versions of French of the systematic initial syllable shortening that was found with the same subjects in the English reiterant productions, although some individual subjects continued to show such a pattern.

Thus, the American teachers of French produced reiterant timing patterns that, while not identical to those of the native French speakers, did not differ significantly from them. On the other hand, the American teachers of French and the French natives both produced final vowel timing patterns that were significantly different from those of the other Americans.

G. General Discussion

There is a growing body of acoustic-phonetic literature that suggests that the non-native productions of late second language learners are influenced, sometimes in subtle ways, by their native language speech patterns (see Flege, 1986, for a review). Most of the research has focused on the analysis of the phonetic characteristics of bilingual speech. Thus the influence of native language phonetic habits has been demonstrated for voice onset time (VOT) in stop consonants for English/French bilinguals (Flege & Hillenbrand, 1984) and for Arabic/English bilinguals (Flege & Port, 1981), because bilinguals show a range of VOT values when speaking their second language that are intermediate between the values produced by monolingual native speakers of the two languages. Native language influences have also been shown for English vowel durations that depend on the voicing of the final consonant, because French/English bilinguals showed vowel durations, when speaking English, that were closer to those of French monolinguals (which vary less with respect to the voicing of a syllable-final consonant) than to those of English-speaking monolinguals (Mack, 1982).

A similar effect of the rhythmic pattern of the native language on the acquisition of the rhythmic patterns of English by native speakers of French has been found by Wenk (1985) who has described his subjects as passing through a transitional "interlanguage" phase, characterized by features of both language systems. Intermediate-level speakers of French who were learning English apparently mastered post-tonic reduced vowels (as in *matter*) before pre-tonic reduced vowels (as in *Japan*), when their productions of such words was judged by native speakers of English. In the present study, native speakers of English who have studied French appear to master the relatively equal durations of nonfinal syllables in French before they master the appropriate French final syllable length, because both groups of American subjects produced essentially equal nonfinal reiterant syllables in French, but only the more experienced group of American subjects, the teachers of French, also produced French-like final syllables. Flege (e.g., Flege, 1981; Flege 1987; Flege & Hillenbrand, 1984) has hypothesized that second language learners may acquire more rapid, accurate pronunciation of a sound that is totally foreign to their native repertoire, because they are unable to assimilate it to one of their native phonemes. Equally-timed nonfinal syllables are not typical of English words, whereas final-syllable stress does occur. Perhaps native speakers of English who learn French are more successful in producing essentially equal nonfinal syllables in their reiterant versions of French than in producing the correct final-syllable lengthening,

because the former pattern is more foreign to their native repertoire.

Many have argued that language learners who begin their study of a second language relatively late fail to master fully the phonetic details of that second language because of biological limitations imposed by a critical or sensitive period for speech acquisition (Lenneberg, 1967; Long, 1990; Oyama, 1979; Scovel, 1988). The notion of a critical period for language acquisition is a strong one and describes a period that is genetically determined, clearly delimited, and not susceptible to the influence of the environment. The notion of a sensitive period for language acquisition, on the other hand, while still a maturational effect, is subject to greater variability, including a less clearly delimited time-frame. Although for some researchers in the field, the onset of adolescence (roughly twelve years of age) was seen as the point after which second language learners were likely to speak their non-native language with a notable foreign accent, others have pushed for acquisition of a foreign accent to six, at least for some individuals (see Long, 1990, for a review). Indeed, Long (1990) has written:

> Thus, while somewhat weaker than the claim for a critical period for first language learning, the claim for a sensitive period for second language acquisition is still a strong and interesting one. The maturational processes underlying it are held to be universal. Hence, learners who begin a second language after its supposed closure (which will here be claimed to be as early as age 6 for phonology in many individuals and around 15 for morphology and syntax), and who nevertheless attain native-like ability in those areas, will falsify the hypothesis (p. 253).

However, all of the native speakers of English in the present study were late learners of French (beginning in junior high school at the earliest), yet the more experienced group of learners (American teachers of French) produced timing patterns that were not significantly different from those of the native French speakers.

Two possible explanations for this pattern of results can be suggested. Either the acquisition of second-language rhythm patterns is exempt from the sensitive period constraint or factors such as length of exposure, training, language aptitude, or motivation may play an important role. Whereas there has been little empirical investigation of the first hypothesis, the role of experience and training has been supported by a number of studies. For example, Wenk (1985) found that his advanced French students of English, unlike those at the intermediate level, had mastered the vowel reduction patterns associated with English word stress. Similarly, Flege and Eefting (1987) found that Dutch speakers of English who majored in the subject were judged to have significantly better pronunciation scores than Dutch students of English who studied to become engineers, although both groups' productions were judged to be significantly different from those of native English speakers. As in the present study, however, experience may have been confounded with aptitude. The English majors, like the university-level teachers of French in the present study, were more experienced second-language learners, but they also probably had greater aptitude for second-language learning. In fact, aptitude rather than experience may be the source of the performance of the group of French teachers. However, in either case, if good reasons for exempting the acquisition of second-language rhythm patterns from the sensitive period constraint are not found, then these results call into question the notion of a sensitive period as currently formulated.

Future research needs to compare directly second-language segmental and rhythmic learning, to see if rhythmic patterns are easier to acquire, and to determine the relative contribution of rhythmic and phonetic factors to the detection of non-native pronunciation.

REFERENCES

Allen, G. D. (1983). Some suprasegmental contours in French two-year-old children's speech. *Phonetica, 40*, 269-292.

Benguerel, A. (1971). Duration of French vowels in unemphatic stress. *Language & Speech, 14*, 383-391.

Clumeck, H. A. (1975). Cross-linguistic investigation of vowel nasalization: An instrumental study. In *Nasalfest: Papers from a Symposium on Nasals and Nasalization* (pp. 133-151). Stanford: Stanford University Press.

Crompton, A. (1980). Timing patterns in French. *Phonetica, 37*, 205-234.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *IRAL, 4*, 183-198.

Di Cristo, A. (1980). La Durée intrinsèque des voyelles du français. *Travaux de l'Institut Phonétique d'Aix, 7*, 211-35.

Duez, D., & Nishinuma, Y. (1985). Le rythme en français: alternance des durées syllabiques. *Travaux de l'Institut de Phonétique d'Aix, 10*, 151-169.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly, 15*, 443-455.

Flege, J. E. (1986). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders Vol. II* (pp. 224-401). Norwood, NJ: Ablex.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. (1988). Factors affecting degree of perceived accent in English sentences. *Journal of the Acoustical Society of America, 84*, 70-79.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*, 185-202.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*, 708-21.

Flege, J. E., & Port, R. (1981) Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*, 125-146.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics, 19*, 193-212.

Hoequist, C. (1983). Syllable duration in stress-, syllable-, and Mora-timed languages. *Phonetica, 40*, 203-37.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208-1221.

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly, 13*, 263-167.

Larkey, L. (1983). Reiterant speech: An acoustic and perceptual validation. *Journal of the Acoustical Society of America, 73*, 1337-45.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Liberman, M. Y., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America, 63*, 231-33.

Lindblom, B. E .F., & Rapp, K. (1973) Some temporal regularities of spoken Swedish. *Quarterly Progress Report of the Speech Transmission Laboratory*, (Stockholm, Sweden: Royal Institute of Technology) 2, 1-12.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*, 251-285.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America, 71*, 173-178.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing "Words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234-45.

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38*, 84-106.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54*, 1235-47.

O'Shaughnessy, D. A (1984). Multispeaker analysis of durations in read French paragraphs. *Journal of the Acoustical Society of America, 76*, 1664-1672.

Oyama, S. C. (1976). A sensitive period for the acquisition of a phonological system. *Journal of Psycholinguistic Research, 5*, 261-283.

Parmenter, C. E., & Blanc, A. V. (1933). An Experimental Study of Accent in French and English. *MLA Publications, 48*, 598-607.

Pike, K. (1945). Intonation of American English. Ann Arbor: University of Michigan Press.

Scovel, T. (1988). *A time to Speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Smith, A. (1977). The timing of French, with reflections on syllable timing. *Edinborough Work in Progress, 10*, 97-107.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53-66). New York: Springer-Verlag.

Vatikiotis-Bateson, E. (1986). *Dynamic analysis of reiterant speech production*. Unpublished doctoral dissertation, Indiana University.

Wenk, B. J. (1985). Speech rhythms and second language acquisition. *Language and Speech, 28*, 157-175.

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10*, 193-216.

FOOTNOTES

Journal of the Acoustical Society of America, 90(6), 3008-3018 (1991).

†Also Wellesley College.

[1] All ratios reported in the paper are to 1.

[2] Results of these analyses of variance were essentially the same, even when all ten original subjects were included. The only significant effect was that of syllable position [$F(1,9)=121.16$, $p<.0000$]. None of the other effects were significant.

[3] In the case of five-syllable words, there were actually four words representing one of the five-syllable word types and two words representing the other.)

[4] For comparability with Oller (1973) secondary stress syllables were grouped with unstressed syllables.

[5] The syllables were here divided into those with primary, secondary and no stress for comparability with Nakatani et al. (1981). The two initial syllables of the second set of five syllable words had complementary stress patterns (one of the words had a secondary stress where the other had no stress and vice versa), so the averaged durations of those syllables were excluded from these calculations.

[6] The results of this analysis and all subsequent analyses include all of the original subjects from both groups. Similar analyses including only the subjects who produced the most consistent reiterant speech produced essentially the same results.

[7] However, the present data exhibit a consistent effect of initial syllable shortening (see Figure 2), which disagrees with findings by Oller (1973), Klatt (1976) and Nakatani et al. (1981). The most likely explanation for this discrepancy is that the reiterant productions in this study were produced as citation forms, rather than in a sentence frame. The present study used citation forms in order to reduce the number of syllables that subjects needed to remember for the reiterant production of individual words (but cf. Nakatani et al., 1981 for a different method). It may be the case that the sentence frame gives extra prominence to the word to be imitated and that such prominence results in the pattern of word-initial syllable length found in the other studies.

APPENDIX

	French Words	English Words (Stress Pattern)
Two syllables	comptoir	counter (SW)
	sacré	sacred (SW)
	progrès	progress (SW)
	contrôle	control (WS)
	surprise	surprise (WS)
	degré	degree (WS)
Three syllables	compliment	compliment (SWW)
	instrument	instrument (SWW)
	solitude	solitude (SWW)
	ingénieur	engineer (WWS)
	indiscret	indiscrete (WWS)
	japonais	japanese (WWS)
	conclusion	conclusion (WSW)
	instructif	instructive (WSW)
	solution	solution (WSW)
	commentaire	
	légendaire	
	société	
Four syllables		commentary (SWWW)
		legendary (SWWW)
	télévision	television (SWWW)
		society (WSWW)
	économie	economy (WSWW)
	publicité	publicity (WSWW)
	exposition	exposition (WWSW)
	population	population (WWSW)
	satisfaction	satisfaction (WWSW)
Five syllables	automatiquemen	automatically (WWSWW)
	élasticité	elasticity (WWSWW)
	électricité	electricity (WWSWW)
	possibilité	possibility (WWSWW)
	communication	communication (WWWSW)
	civilisation	civilization (WWWSW)

(S=primary stress, W=secondary stress or no stress)

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 73-88

Syllable-internal Structure and the Sonority Hierarchy: Differential Evidence from Lexical Decision, Naming, and Reading*

Andrea Levitt,† Alice F. Healy,†† and David W. Fendrich†††

Treiman (e.g., 1983) and others have argued that spoken syllables are best characterized, not as linear strings of phonemes, but as hierarchically organized units consisting of an onset (initial consonant or consonant cluster) and a rime (the vowel and any following consonants) and that the rime is further divided into a peak or nucleus (the vowel) and a coda (the final consonants). It has also been argued that the sonority (or vowel-likeness) of the consonant closest to the peak, which is a function of its phonetic class, may have an effect on the strength of boundaries determined by the hierarchical division of the syllable (e.g., Treiman, 1984). We examined the evidence for syllable-internal structure and for sonority in two experiments that employed *visually presented* stimuli and lexical decision, naming, and reading tasks. Our results provide support for the breakdown of the rime into a peak and a coda and for an effect of the sonority of the postvocalic consonant on that break. This pattern occurred only in our lexical decision tasks, so the effect is assumed to be postlexical. We did not find an effect of the onset-rime boundary, perhaps because of an unanticipated effect of word frequency. Our results are discussed in terms of phonological coding in short-term memory.

Recent psycholinguistic evidence has suggested that English syllables are organized hierarchically, divided first into an *onset* (consisting of the initial consonant or consonant cluster) and a *rime* (consisting of the following vowel and any additional consonants), with the rime further divided into a *peak* or *nucleus* (consisting of the vowel) and a *coda* (consisting of the remaining consonants).[1] For example, Cooper, Whalen, and Fowler (1986) have shown that the P-center (moment of perceptual occurrence) of a syllable depends on the duration, though not the number, of syllable initial consonants (the onset) and, in a later study, to a lesser extent on the rime (Cooper, Whalen, & Fowler, 1988). This division of the syllable into an onset and rime is particularly well supported by a number of studies by Treiman (1983, 1986), who taught subjects novel word games in which they were required to recombine components from pairs of nonsense syllables or words, and found that they were more likely to divide those syllables between the onset and rime than elsewhere in order to complete the tasks. More recently, Treiman and Chafetz (1987) have demonstrated evidence for the onset/rime break in *printed* words, using both an anagram and a lexical decision task. In the first case, they found that subjects were better able to recognize a word like *twist* when it was divided TW IST (at the onset/rime boundary) than when it was divided TWI ST (between the peak and the coda). In the second case, subjects responded more quickly in a lexical decision task when the test item contained slashes after the onset (CR//ISP) than after the vowel (CRI//SP).

Preparation of this article was supported in part by NICHD grant HD-01994 to Haskins Laboratories and by United States Army Research Institute Contract Number MDA903-86-K-0155 to the Institute of Cognitive Science at the University of Colorado.

We thank Carol Fowler and Rebecca Treiman for their comments on earlier versions of the manuscript.

We gratefully acknowledge the assistance of Antoinette Gesi and Miriam Butt in the data collection.

The evidence in support of dividing the rime into a nucleus and a coda is perhaps somewhat less compelling. Treiman (1983), using novel word games, found only weak support for the nucleus/coda division and suggested that the division might depend on the phonetic makeup of the final consonant cluster. Indeed, when she systematically varied the sonority (or vowel-likeness) of the consonant following the vowel in VCC syllables (Treiman, 1984), she found that subjects in a word game task tended to view liquid consonants, which are quite vowel-like, as belonging to the nucleus or peak, obstruents, which are not at all vowel-like, as belonging to the final consonant cluster or coda, and nasals, which are intermediate in terms of sonority, as showing an equal affinity to both the nucleus and the coda. Derwing, Nearey, and Dow (1987) obtained similar results. These findings are largely in agreement with the proposals of MacKay (1972) and Stemberger (1983) that liquids following the vowel be assigned to the nucleus rather than the coda. The findings also agree with the sonority hierarchy proposed for syllables (e.g., Hooper, 1976), which suggests that syllable peaks are peaks of sonority, that consonant classes vary with respect to their degree of sonority, or vowel-likeness, and that segments on either side of the peak show a decrease in sonority with respect to the peak.

However, the evidence connecting the ease of the onset-nucleus break to the sonority of the *pre*vocalic consonants has been less consistent. Treiman (1986) found that there was no effect of the phonetic category of the *pre*vocalic consonant on the onset-rime division (suggesting that onsets consisting of more than one consonant remain cohesive), while Derwing et al. (1987) did find such an effect of the phonetic category of the prevocalic consonant.

Most of the evidence for the hierarchical division of the syllable into an onset and rime, and possibly into a nucleus and coda, comes from studies that present stimuli auditorily and require subjects to focus closely on the phonological structure of the stimuli in order to play novel word games or perform segment interchanges. The literature on reading is divided as to whether the phonological code of a visual stimulus is obligatorily accessed (see, e.g., Van Orden (in press)) or whether it is accessed only under certain circumstances (e.g., McCusker, Hillinger, and Bias (1981)). One study that used visual stimuli and looked for evidence of the hierarchical division of the syllable was done by Treiman and Chafetz (1987). As mentioned above, they required subjects to perform either an anagram or a lexical decision task on visually presented stimuli, however, they only compared subjects' responses to stimuli with breaks between the onset and the rime with their responses to stimuli with breaks following the nucleus. They did not examine the effects of breaks within initial and final consonant clusters as compared to the two breaks mentioned above, nor did they investigate, in this study, the effect of sonority on the strength of these divisions. As a result of her numerous studies, Treiman (1986) has suggested that the intrasyllabic organization of the syllable should be recognized in theories of speech perception and production as well as in theories of reading.

Research that has compared the results of lexical decision and word naming tasks (e.g., Seidenberg, Waters, Sanders, & Langer, 1984) suggests that certain effects may be postlexical, i.e., a result of processing that occurs after lexical access. Thus, such effects emerge only in lexical decision and not in naming tasks, since naming typically takes less time and is thus believed to involve less postlexical processing. It is often assumed, however, that naming a visually presented word requires accessing its phonological code (e.g., Seidenberg, 1985). Silent reading of visually presented stimuli is another task that has been shown to be sensitive to semantic and phonological priming (McNamara & Healy, 1988), while also presumably requiring less postlexical processing. It would be of interest, therefore, to see whether evidence for the hierarchical structure of the syllable can be found in each of these three tasks.

The present experiments are thus designed (a) to replicate Treiman's (1984) finding that the break between the nucleus and coda varies as a function of the phonetic class (liquid, nasal, or obstruent) of the postvocalic consonant, with postvocalic liquids showing the greatest cohesion to the nucleus and obstruents showing the least, (b) to test for a similar effect of the phonetic class of the prevocalic consonant on the break between the onset and the rime, and (c) to determine whether any evidence for such breaks is pre- or postlexical in origin by comparing the results of lexical decision tests with those of naming and reading.

EXPERIMENT 1

Subjects responded orally to visually presented stimuli, including both words and nonwords, all of which were monosyllabic and five letters long. Each visually presented stimulus could be interrupted at one of six possible locations by an

asterisk. One group of subjects performed a lexical decision task while a second group named each of the items out loud.

Method

Stimuli. Two sets of test items were constructed, one to examine the effect of the composition of initial consonant clusters on the cohesion of the onset-rime boundary of the syllable and another to examine the effect of the composition of final clusters on the cohesion of the rime-internal nucleus-coda boundary. All test items were single syllables, contained five letters, and, with the exception of some of the onset-rime test items, described below, all had a $C_1C_2VC_3C_4$ phonemic structure.

In the case of the onset-rime test words, C_2 was either a liquid (twelve items), a nasal (six items) or an obstruent (six items).[2] There were twelve additional five-letter words with no initial consonant cluster, but with an initial single phoneme, e.g., /ʃ/, which is normally represented by two letters, "sh."[3] Nine of these items had a $C_1VC_2C_3$ phonemic structure, and three had a C_1VC_2 structure. All were five letters long. All words were also low frequency, with the mean frequency for the liquid items 7.3 (Kučera & Francis, 1967), for nasal items 9.8, for obstruent items 6.8, and for single-phoneme items 7.8. The corresponding onset-rime nonword test items were constructed by switching the vowel and final consonants of one item with the vowel and final consonant of another item from the same series, so that two nonwords were created (e.g., *craft* and *flint* giving *flaft* and *crint*).

In the case of the nucleus-coda test words, there were twelve words each for which C_3 was a liquid, a nasal, or an obstruent. The corresponding nucleus-coda nonword test items were constructed as above (e.g., *blunt* and *swamp* yielding *swunt* and *blamp*). The mean frequency for the liquid items was 9.8, for the nasal items 9.1, and for the obstruent items 9.8.

Each word and nonword (see Appendix for the complete list) could appear with an asterisk in one of three positions. For the onset-rime test items, the asterisk could appear before the word (Position 1), after the first letter (Position 2), or between the second letter and the vowel (Position 3), e.g., *CRAFT, C*RAFT or CR*AFT. For the nucleus-coda test items, the asterisk could appear after the vowel (Position 4), after the third consonant (Position 5), or after the word (Position 6), e.g., BLU*NT, BLUN*T, BLUNT*. Positions 1 and 6 are control positions because the asterisk does not interrupt either the initial or the final consonant cluster.

Three lists of 144 test items were prepared. Each word and nonword appeared only once on each list.[4] The order of presentation was pseudorandom with the following constraints: In every twelve items there was an equal number of onset-rime and nucleus-coda test words and nonwords and an equal number of asterisks at each of the six positions. For the nucleus-coda test items, in every group of twelve, there were two stimuli with a liquid, nasal, or obstruent as the C_3 phoneme. For the onset-rime items, in the same group of twelve, there were two stimuli with a liquid as the C_2 phoneme, two stimuli with a single initial consonant, and either two stimuli with a nasal as the C_2 phoneme or two stimuli with an obstruent as the C_2 stimuli. The three lists differed only as to the location of the asterisks with each one of three possible asterisk locations occurring once across lists for every stimulus.

Procedure. Subjects were told that strings of letters would appear on the computer screen in front of them. Subjects in the lexical decision condition were to say "yes" if the string was a word and "no" if it was not. Subjects in the naming condition were to read the word or nonword out loud. A voice key was used to record subjects' response times. The experimenter first made sure that the key was responding properly to the level of the subject's voice, and the subject was instructed not to make inadvertent noises, as the key was quite sensitive. Subjects' responses were recorded on cassette tapes. The experimenter noted all errors in both conditions, so that the responses to those items would be excluded from analysis.

Subjects. Twenty-four Wellesley College undergraduates were paid for their participation in the experiment and were assigned to conditions by order of arrival, according to a fixed rotation.

Results

The onset-rime and nucleus-coda words represented different sets of words[5] and therefore each set of items was analyzed separately. All response latencies were reciprocally converted to speeds for the analyses,[6] but the resulting mean speeds were converted back to latencies for reporting in the text and in the figures. Two sets of analyses were performed, one on the latencies for correct responses and another on the error proportions. A response was considered an error in the naming task if a subject failed to respond or if the response was incorrect. Items were not treated as a random effect because the stimuli were not randomly selected (Wike & Church, 1976).

We need to obtain an effect of asterisk position in order to demonstrate syllable-internal structure

and an interaction of asterisk position with cluster composition in order to demonstrate an effect of the sonority hierarchy on syllable-internal cohesiveness.

Onset-rime words. For the onset-rime test words, the analyses included one between-subjects factor, response condition (lexical decision or naming) and two within-subjects factors, onset composition (C_2 either an obstruent, liquid, nasal, or the second grapheme of a single phoneme) and asterisk position [immediately preceding the word (Position 1) or following the first (Position 2) or second (Position 3) letter]. In these analyses we did not find the anticipated asterisk position effect nor the anticipated asterisk position by onset composition interaction. However, we did find some interesting effects of response condition and onset composition.

As would be expected if, indeed, the lexical decision task requires additional postlexical processing, the mean latency for naming (673 ms) was faster than that for lexical decision (831 ms). Likewise, the error proportions were higher for lexical decision (.092) than for naming (.042). In the overall analysis of response latency for onset-rime words, there was a significant main effect of response condition (lexical decision vs. naming), $F(1,22)=12.37$, $p=.0022$, MSe=5.7176. The effect of response condition was also significant in the error analysis, $F(1,22)=8.36$, $p=.0083$, MSe=.1824.

Although the differences in frequency were small among the words comprising the different onset composition groups, the differences in frequency seem to have produced corresponding differences in both mean latencies and error proportions. Recall that the mean frequency for nasals was 9.8, for one-phoneme items 7.8, for liquids 7.3, and for obstruents 6.8. Correspondingly, the mean latencies for nasal test items (722 ms),one-phoneme items (729 ms), liquids (734 ms), and obstruents (796 ms) increased as the items became less frequent, as did the error proportions, with one small reversal (nasal=.028, one-phoneme=.067, liquid=.061, obstruent=.111). In the latency analysis, there was a significant main effect of onset composition, $F(3,66)=6.26$, $p=.0011$, MSe = .2513, which was also significant in the error analysis, $F(3,66)=4.33$, $p=.0077$, MSe=.0844.

The effect of onset composition, which presumably reflected the frequency of the words for each of the onset composition types, was evident for the lexical decision task but not for the naming task. As was the case for the combined data, for the data from the lexical decision task, latencies increased as word frequency declined, and error proportions also increased, with one small reversal. The latencies and error proportions in the lexical decision task were 766 ms and .042 for nasals, 808 ms and .093 for one phoneme, 844 ms and .081 for liquids, and 923 ms and .153 for obstruents. There was a significant interaction of onset composition with response condition, $F(3,66)=3.61$, $p=.0174$, MSe=.1449, in the latency analysis, but not in the error analysis. In a separate planned analysis of the lexical decision latencies, done to investigate the source of this interaction, there was a significant effect of onset composition, $F(3,33)=6.85$, $p=.0013$, MSe=.3114, which was marginally significant as well in an error analysis of the lexical decision task, $F(3,33)=2.57$, $p=.0699$, MSe=.0762. There were no significant effects in either the latency or the error analysis of the naming data, so this pattern seems limited to the lexical decision data (see Figure 1).

Nucleus-coda words. For the nucleus-coda test words, the analyses included one between-subjects factor, response condition (lexical decision or naming) and two within-subjects factors, coda composition (C_3 either an obstruent, liquid, or nasal) and asterisk position [immediately following the vowel (Position 4) or following C_3 (Position 5) or C_4 (Position 6)].

As found for the onset-rime words and as expected under the assumption that the lexical decision task requires more (postlexical) processing than does the naming task, the mean latency for lexical decision (830 ms) was longer than for naming (657 ms). Likewise, the mean error proportion for lexical decision was higher (.103) than for naming (.029). In the analysis of the nucleus-coda test items, there was a significant effect of response condition, lexical decision vs. naming, $F(1,22)=15.26$, $p=.0010$, MSe=5.3887. This effect was also significant in the overall error analysis, $F(1,22)=22.78$, $p=.0002$, MSe=.3025.

Just as there was an effect of onset composition for the onset-rime words, there was an effect of coda composition for the nucleus-coda words. However, the effect in this case was only evident for latencies, not errors, and did not reflect differences in word frequency, which were minimal. The overall latency (combining data from the lexical decision and naming tasks) to C_3 obstruents (708 ms) was shorter than to nasals (723 ms), which were in turn shorter than to liquids (775 ms). (See Figure 2.) There was a significant main effect in the latency data of the coda composition, $F(2,44)=16.66$, $p<.0001$, MSe=.2914, not significant in the error analysis.

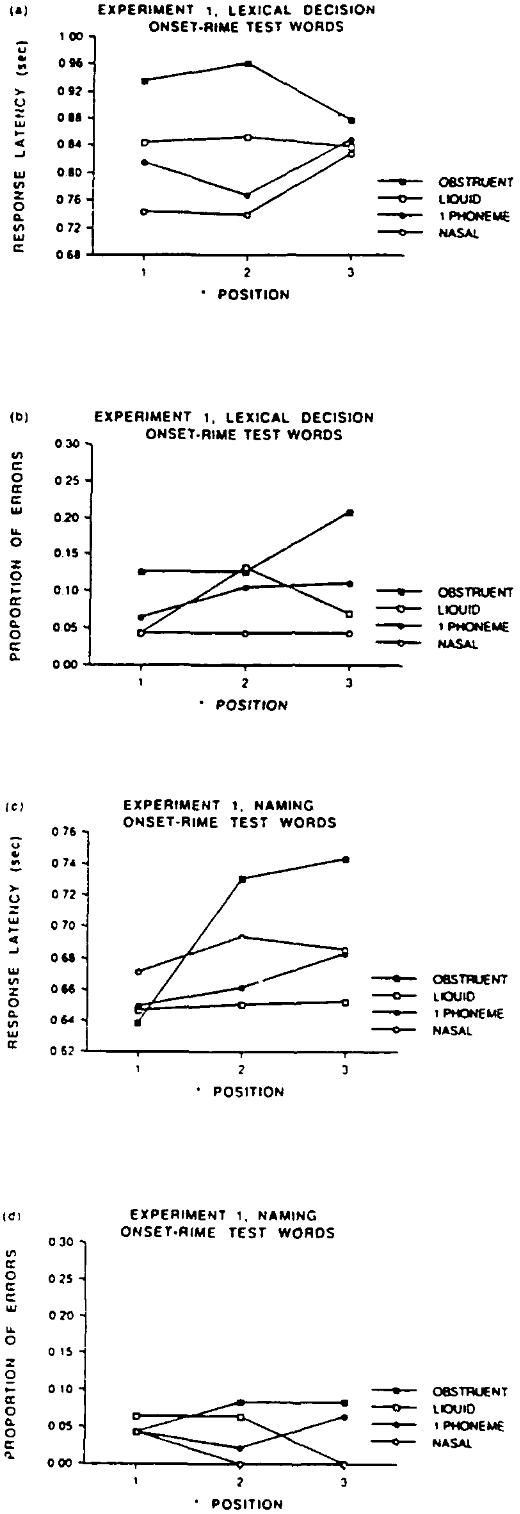

Figure 1. **The results of the onset-rime test words in Experiment 1 as a function of the phonetic class of C_2 and of asterisk position. The asterisk appears before the word at Position 1, after the first letter at Position 2, and between the first and second letter at Position 3. Panels (a) and (b) are for the lexical decision task; panels nd (d) are for the naming task. The latency analysis is shown in panels (a) and (c); the error analysis is shown in pa (b) and (d).**

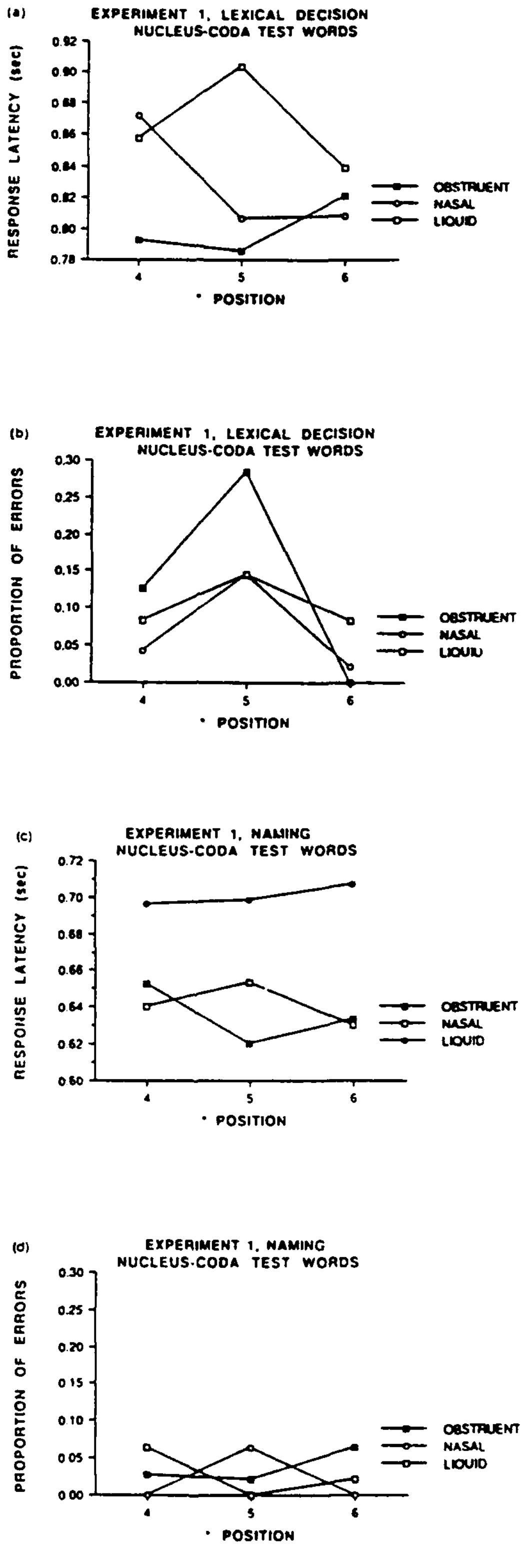

Figure 2. **The results of the nucleus-coda test words in Experiment 1 as a function of the phonetic class of C_3 and of asterisk position. The asterisk appears after the vowel at Position 4, between the last two letters at Position 5, and after the word at Position 6. Panels (a) and (b) are for the lexical decision task; panels (c) and (d) are for the naming task. The latency analysis is shown in panels (a) and (c); the error analysis is shown in panels (b) and (d).**

None of the remaining effects in the latency analysis were significant. Most crucially, there was no effect of asterisk position or interaction of asterisk position and coda composition. There were, however, several other interesting effects in the error analysis, and the expected effect of asterisk position and the expected interaction of asterisk position and coda composition were evident for the lexical decision task, but not for the naming task. (See Figure 2.) Overall error proportions in Position 5 (.110) were higher than in Position 4 (.057) or in Position 6 (.031). Whereas the most errors occurred in Position 5 overall and for all coda compositions with the lexical decision data, with the naming data the most errors occurred in Position 6 for the obstruents, in Position 5 for the nasals, and in Position 4 for the liquids. (See Figure 2.) The main effect of asterisk position was significant in the error analysis, $F(2,44)=10.42, p=.0004$, MSe=.1161. There was also an asterisk position by response condition interaction, $F(2,44)=10.57$, $p=.0004$, MSe=.1178, and a three-way interaction of asterisk position by response condition by coda composition, $F(4,88)=2.92, p=.0253$, MSe=.0362.

As with the onset-rime words, planned analyses were conducted on the data with the nucleus-coda words separately for the lexical decision and naming tasks. For the lexical decision latencies, as for the combined latencies, there was an effect of coda composition, with latencies to items in which C_3 was an obstruent shorter (799 ms) than those in which C_3 was a nasal (828 ms), which were in turn shorter than those in which C_3 was a liquid (866 ms). There was a significant main effect of coda composition in the analysis of the lexical decision latencies, $F(2,22)=5.44$, $p=.0120$, MSe=.0841.

For the lexical decision errors, as for the combined errors, there was an effect of asterisk position, with the most errors (.19) occurring when the asterisk appeared between C_3 and C_4 in Position 5, next most (.08) when the asterisk appeared between the vowel and C_3 in Position 4, and fewest (.04) when the asterisk appeared at the end of the word in Position 6. However, as anticipated, the effect of asterisk position depended on coda composition to some extent. As can be seen in Figure 2, obstruents and, to a lesser extent, nasals showed a dramatic increase in the proportion of errors when the asterisk intervened at Position 5 between C_3 and C_4, but the increase in errors for liquid items with an asterisk at Position 5 was less pronounced. There was a significant main effect of asterisk position in the error analysis of the lexical decision task, $F(2,22)=14.50$, $p=.0002$, MSe=.2339. There was also a marginally significant interaction of asterisk position and coda composition, $F(4,44)=2.44, p=.0601$, MSe=.0400.

For the naming latencies, as for the lexical decision latencies and the combined latencies, responses to items with C_3 as an obstruent were shorter (635 ms) than to those with a nasal (641 ms), which in turn were shorter than to those with a liquid (700 ms). The main effect of coda composition was significant, $F(2,22)=12.06$, $p=.0004$, MSe=.2355, and there were no other significant effects in the analysis of naming latencies. There were no significant effects at all in the error analysis of the naming data.

Discussion

Our analysis of the words designed to test the cohesiveness of the onset-rime boundary and the possible effect of the sonority hierarchy on that boundary produced some surprising results. There were no effects of syllable-internal structure or sonority in the naming data. The lexical decision data also failed to demonstrate any such effects, but showed an apparent effect of word frequency, in both the latency and error analyses.

The analysis of the nucleus-coda test items proved somewhat more promising with respect to syllable structure and sonority (see also Treiman, 1984, 1986). In the overall error analysis, there were significantly more errors when the asterisk intervened at Position 5 (between the two consonants of the coda) than at Position 4 (immediately after the vowel) or at Position 6 (at the end of the word). These results suggest that interruption at the nucleus-coda boundary (after the vowel) is less disruptive than within the coda itself. In both the separate lexical decision and naming latency analyses there were significant main effects of coda composition, with responses to obstruent items faster than to those with a nasal, which in turn were faster than those with a liquid. Indeed, this was the only significant effect found in the separate analysis of the naming data. On the other hand, in the error analysis of the lexical decision data, there was a significant effect of asterisk position, showing that the disruptive effect of the asterisk appearing within the coda is a postlexical effect. Finally, there was also a marginally significant interaction for the lexical decision error analysis of the coda composition with asterisk position. This interaction provided partial support for the notion that the class of the *post*vocalic consonant affects the cohesiveness of

the nucleus and the coda. Postvocalic obstruents are lowest on the sonority hierarchy. Thus, they are expected to show the *least* cohesiveness with the preceding vowel and the *most* cohesiveness with the final consonant, followed by nasals and then liquids. As Figure 2 illustrates, errors were greatest for test items with an obstruent when the asterisk interrupted the rime at Position 5. Nasal test items showed a similar disruption in that position. On the other hand, liquid test items should have shown more errors with an asterisk in Position 4, rather than an increase at Position 5, because of the greater cohesiveness of liquids to the preceding vowel. However, that was not the case.

Because of the constraints we followed in constructing the stimuli for this experiment, it was not possible to have all test items begin with the same sound, which would have been ideal since we used a voice key to record subjects' responses. We wondered whether the various phonetic identities of the first consonants of our test items had had an effect on the naming speeds. We also wondered whether our use of the voice key to record subjects' responses in the lexical decision task had introduced greater variability in the response times than would have been the case with a reaction time key (see, e.g., Pechmann, Reetz, & Zerbst, 1989). If so, it might explain why our evidence for syllable-internal structure and for some influence of the sonority hierarchy on the nucleus-coda boundary only emerged in the error analysis. We decided to repeat the experiment using a manual reaction time key and substituting a silent reading task for our naming task.

EXPERIMENT 2

In this experiment, we compared the responses of one group of subjects in a lexical decision task to those of another group of subjects whose task was to read the word and nonword stimuli silently and to press a key as soon as they were done with each item. McNamara and Healy (1988) have demonstrated semantic and rhyme facilitation with a self-paced reading task of this type, which, however, like naming, is assumed to involve less postlexical processing than lexical decision.

Method

Stimuli. The same stimuli used in Experiment 1 were used in Experiment 2.

Procedure. The procedure was essentially the same as in Experiment 1, except that a reaction time key was used instead of a voice key. Subjects in the reading condition were to read the word or nonword silently and to press a button with the index finger of their right hand as soon as they had finished reading each item. Subjects in the lexical decision condition were to decide whether or not each letter string was an English word. They were told to rest the index finger of their right hand on the "yes" button and the index finger of their left hand on the "no" button, and to press "yes" as quickly as possible if the string was a word, and "no" as quickly as possible if the string was not a word. They were told that both speed and accuracy would be scored by the computer.

Subjects. Thirty-six male and female undergraduate students from the University of Colorado at Boulder participated in this experiment. They received course credit for their participation. They were assigned to conditions by order of arrival, according to a fixed rotation.

Results

As in Experiment 1, the onset-rime and nucleus-coda test words were analyzed separately. Two sets of analyses were performed, one on error rates (for the lexical decision data only) and another on the latencies for correct responses. All response latencies were reciprocally transformed to speeds for the analyses, but the resulting mean speeds were converted back to latencies for reporting in the text and in the figures, as for Experiment 1. Also as for Experiment 1, items were not treated as a random effect because the stimuli were not randomly selected (Wike & Church, 1976).

Onset-rime words. As in Experiment 1 and as anticipated given that the lexical decision task presumably requires additional postlexical processes not included in the reading task, the mean latency for reading (758 ms) was considerably shorter than that for lexical decision (920 ms). In the overall latency analysis of the onset-rime test items there was a significant main effect of lexical decision vs. reading, $F(1,34)=4.56$, $p=.0378$, MSe=5.8280. Also in accord with predictions, based on the assumption that the asterisk should be least disruptive when it precedes the word, the response latency for stimuli with immediately preceding asterisks (Position 1) was shorter than for stimuli with asterisks in Positions 2 and 3 (1.234 vs. 1.188 and 1.189, respectively). There was a significant main effect for asterisk position, $F(2,68)=4.44$, $p=.0152$, MSe= .0999.

Also as in Experiment 1, despite the small differences in frequency among the words

comprising the different onset composition groups, the average latency for each onset composition varied largely as a function of the frequency of the words in the four groups, with more frequent words producing shorter latencies. Thus, latency of response (805 ms) was shortest to nasal test items (mean frequency 9.8), followed by the latency of response (817 ms) to single-phoneme test items (mean frequency 7.8), followed by a minor reversal, with response latency (855 ms) to liquid test items (mean frequency 7.3) slightly slower than average latency (850 ms) to obstruent test items (mean frequency 6.8). There was a main effect of onset composition, $F(3,102)=8.43$, $p=.0001$, MSe=.1338, as well as a significant interaction of onset composition with lexical decision vs. reading, $F(3,102)=5.21$, $p=.0026$, MSe=.0826.

Separate planned analyses of the reading and lexical decision onset-rime word data were conducted to explore the source of the interaction. In Experiment 1 the correlation of word frequency and onset composition class was evident for the lexical decision task but not for the naming task. Similarly, the correlation of word frequency and onset composition class in the present experiment occurred in the lexical decision task but not in the reading task. There was an effect of onset composition on reading, but this effect was clearly due to the difference between those words in which C_2 was a liquid (777 ms) and all the others (obstruent = 751 ms, one phoneme = 751 ms, and nasal = 752 ms). In the separate reading analysis, there was a significant effect of onset composition, $F(3,51)=3.40$, $p=.0241$, MSe=.0254. There were no other significant effects in the reading analysis.

The latency data from the lexical decision task alone mirror the combined data from both tasks. As in the overall data, latencies in the lexical decision task to words where the asterisk appeared at the beginning were faster (884 ms) than those to words where the asterisk appeared after the initial consonant (940 ms) or just before the vowel (937 ms), as expected because the asterisk should be more disruptive when it occurs in the middle of a word than when it precedes the word. The effect of asterisk position was marginally significant in the lexical decision latency analysis, $F(2,34)=3.16$, $p=.0538$, MSe=.1022.

As can be seen in Figure 3, both the pattern of errors (which were analyzed for the lexical decision task only, because no errors were possible in the reading task) and the pattern of response latencies for the lexical decision task varied as a function of the frequency of the four groups of words, with error proportions (with the exception of one small reversal) lower for more frequent items, and with latencies shorter for the more frequent items, as in Experiment 1. The mean latencies, given in terms of nasal, single-phoneme, liquid, and obstruent test items (that is, in order from most to least frequent), were 866 ms, 895 ms, 949 ms, and 977 ms, whereas the error proportions (in the same order) were .046, .120, .111, and .185. There was a significant effect of onset composition for both the latencies, $F(3,51)=7.87$, $p=.0004$, MSe=.1910, and the error proportions, $F(3,51)=5.31$, $p=.0032$, MSe=.1740.

Nucleus-coda words. As found in Experiment 1 and for the onset-rime words in the present experiment and as expected under the assumption that the lexical decision task requires more postlexical processing than does the reading task, the mean overall latency for reading nucleus coda words (765) was considerably shorter than that for lexical decisions on those words (939). For the combined analysis of the nucleus-coda test word latencies, there was a significant main effect of lexical decision vs. reading, $F(1,34)=5.14$, $p=.0281$, MSe=4.751.

Just as we predicted and found that asterisks were less disruptive when they preceded a word than when they occurred in the middle of a word for the onset-rime stimuli, the asterisks should be less disruptive when they follow a word than when they occur in the middle of a word for the nucleus-coda stimuli. Indeed, the latency for words with item-final, Position 6 asterisks (825) were shorter than those for Position 4 asterisks (842), which were in turn shorter than those for Position 5 (862). There was a significant main effect of asterisk position in the combined analysis of the nucleus-coda test word latencies, $F(2,68)=5.30$, $p=.0074$, MSe=.0730.

Most crucial is the predicted interaction of coda composition and asterisk position. The predicted pattern was found for the lexical decision latencies, but not for the errors in the lexical decision task nor for the latencies in the reading task. As anticipated, the obstruents and nasals showed longer lexical decision latencies at Position 5, whereas the liquids showed the longest lexical decision latencies at Position 4. (See Figure 4). In a separate planned analysis of the lexical decision latencies, there was, in addition to a significant main effect of asterisk position, $F(2,34)=5.7$, $p=.0075$, MSe=.0990, a marginally significant interaction of coda composition and asterisk position, $F(4,68)=2.44$, $p=.0543$, MSe=.0335.

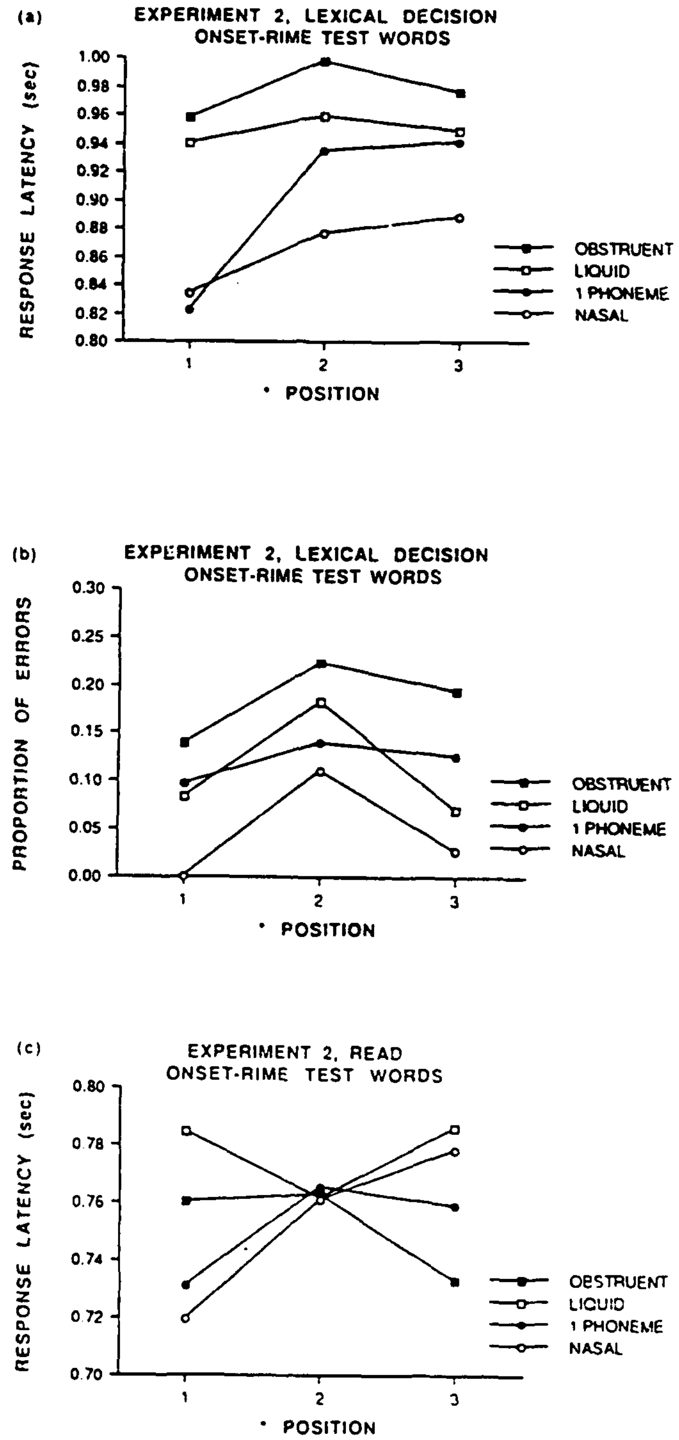

Figure 3. The results of the onset-rime test words in Experiment 2 as a function of the phonetic class of C_2 and of asterisk position. The asterisk appears before the word at Position 1, after the first letter at Position 2, and between the first and second letter at Position 3. Panels (a) and (b) are for the lexical decision task; panel (c) is for the reading task. The latency analysis is shown in panels (a) and (c); the error analysis is shown in panel (b).

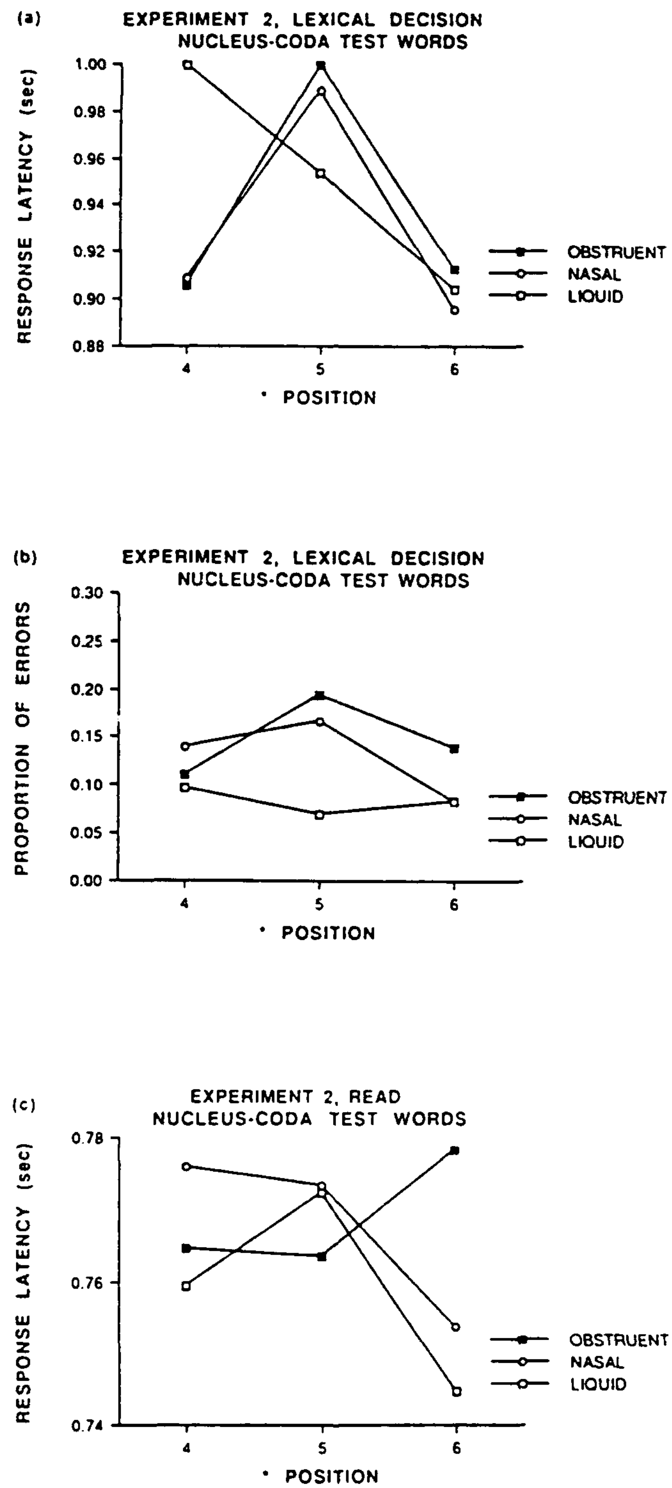

Figure 4. **The results of the nucleus-coda test words in Experiment 2 as a function of the phonetic class of C3 and of asterisk position. The asterisk appears after the vowel at Position 4, between the last two letters at Position 5, and after the word at Position 6. Panels (a) and (b) are for the lexical decision task; panel (c) is for the reading task. The latency analysis is shown in panels (a) and (c); the error analysis is shown in panel (b).**

It should be noted that although this crucial interaction was only marginally significant by this test, the statistic used was very conservative because it was not directional. If a directional test were employed (which seems appropriate in this case because a specific pattern of results was anticipated and obtained), then the results would be clearly significant. In any event, there were no significant effects in the separate analysis of mean proportion errors for lexical decision nor in the separate analysis of the latencies for the reading data.

Discussion

As in Experiment 1, we found highly consistent significant differences between our two tasks in both the latency and error analyses. These significant effects are, of course, consistent with the notion that the lexical decision task requires additional processing.

When we consider the onset-rime data, the most interesting effect that emerged is the effect of onset composition, such that speeds and error rates varied largely as a function of the frequency of the stimuli in each of the onset-composition groups. As in Experiment 1, both the separate latency and error analyses of the lexical decision data showed that responses to the different onset-composition groups varied as a function of their frequency. On the other hand, the main effect of onset composition in the separate latency analysis of the reading data was due to slower response times to stimuli with liquids as the second consonant. There was also an effect of asterisk position in the lexical decision latency analysis, but it provided no support for the internal structure of the syllable, because there was no difference in the latencies to words with asterisks appearing within the onset as compared to those with asterisks between the onset and the vowel. But, the response latencies in both of those positions was marginally significantly slower than when the asterisk appeared at the very beginning of the word.

As in Experiment 1, it was only the analysis of the nucleus-coda data that provided some support for the notion of syllable-internal structure and for the influence of the sonority hierarchy on that structure. Thus, asterisks placed between the nucleus and the coda were less disruptive than those placed within the coda, for the lexical decision analysis. More importantly, in the separate latency analysis of the lexical decision data, there was an interaction (which was marginally significant by a conservative non-directional test) between asterisk position and coda composition, so that test items with postvocalic liquid consonants produced the slowest latency of response when the asterisk appeared immediately after the vowel in Position 4, whereas test items with postvocalic nasals and stops produced the slowest speeds of response when the asterisk appeared just before the final consonant in Position 5. This pattern is consistent with an effect of the sonority hierarchy on the nucleus-coda boundary, because liquids are higher on the sonority hierarchy and therefore more cohesive with the preceding vowel (hence the slower latency for asterisks in Position 4), whereas obstruents and nasals are lower on the sonority hierarchy and therefore more cohesive with the following consonant (hence the slower speeds for asterisks in Position 5).

GENERAL DISCUSSION

We found evidence, but only in our lexical decision tasks, in support of the division of the rime into a nucleus and a coda as well as evidence that suggests that the sonority of the postvocalic consonant affects the strength of that break. It appears from our data that these syllable-structure effects are postlexical (occurring in the lexical decision rather than in the naming or reading tasks).

On the other hand, despite the wealth of psycholinguistic evidence supporting the syllable-internal structures of onset and rime, we were unable to find evidence to support this division in our two experiments. Instead, we found evidence of a word frequency effect, even though we controlled for word frequency,[7] such that the differences among the word frequencies in the four onset groups were not significant. This unanticipated word-frequency finding has potential methodological import. Given multiple experimental constraints, researchers have probably been unable in many cases to find exact frequency matches for their stimuli. They have probably generally assumed that small frequency differences of the type that separated our groups of onset-rime words would be unlikely to produce any effect. Furthermore, the finding also has theoretical import, since these small frequency differences turn out, at least in this case, to matter significantly. Indeed, our word frequency effect was strong enough, occurring in both experiments and for both accuracy and latencies, to override any effect of the onset-rime break.

We would suggest that previous studies that supported the notion of a break between the onset and rime, even with nonword stimuli, were able to find such evidence because the tasks that they employed relied largely on a form of phonological coding used to maintain information in short-term memory, a form of phonological coding which may not be required by simple naming and reading tasks.

Besner and Davelaar (1982) present evidence that the phonological code used to achieve lexical access from print is *not* the same phonological code used to maintain information in short-term memory. In particular, they found that subjects better recalled nonwords with an entry in the phonological lexicon (e.g., BRANE) than nonwords without such an entry (e.g., SLINT) even under conditions of articulatory suppression, whereas effects of phonological similarity and word length were eliminated by articulatory suppression. Because of the opposing effects of articulatory suppression, they argue that there are two phonological codes. The first phonological code permits lexical access, whereas the second code, more strongly affected by articulatory suppression, is used to maintain information in short-term memory. If we assume that the first phonological code not only permits lexical access but also subserves naming and that effects of syllable structure and sonority emerge through use of the second, short-term-memory phonological code, then we can reconcile our results with those of previous studies.

The majority of the psycholinguistic studies finding evidence in support of the hierarchical structure of the syllable involve tasks that require the maintenance of information in short-term memory. The novel word games task used frequently by Treiman (e.g., 1983, 1984, 1986) and the substitution-by-analogy task (where subjects switch specified parts of two jointly presented monosyllabic strings) used by Derwing et al. (1987), Dow (1987), Fowler (1987) and others involve such a demand. Thus, it is reasonable to assume that they required use of the phonological code that maintains information in short-term memory and from which effects of syllable structure and sonority emerge. Indeed, Treiman and Danis (1988) demonstrated syllable structure effects using a short-term memory task.

Perhaps lexical decision, unlike naming and reading, makes a greater demand on short-term memory. For example, subjects in a lexical decision task may store accessed items in short-term memory for decision processing. Our consistently significant differences between lexical decision, on the one hand, and naming and reading on the other, support, as do many other studies, the notion of additional post-lexical processing in lexical decision tasks. We suggest that this processing may entail maintenance of the accessed item in short-term memory. If evidence for the syllable's internal organization and for the influence of the sonority hierarchy on that organization emerges only in tasks that require the maintenance of information in short-term memory, and if lexical decision requires such maintenance, then it is not surprising that our results supporting syllable-internal structure emerged only in the lexical decision task.

However, we found support only for the breakdown of the rime into a peak and a coda, whereas Treiman and Chafetz (1987) found, also using a lexical decision task, that subjects responded more rapidly to visually presented words and nonwords when slashes appeared between the onset and the rime than when they appeared between the peak and the coda. There are at least two possible sources for this discrepancy. In the first place, they compared visual interruptions after the onset and after the peak within the same set of words and nonwords, whereas we used different words to test the strength of the onset-rime boundary and the nucleus-coda boundary. We thus could not compare directly the strength of these two boundaries. Secondly, we found an unanticipated, significant effect of onset type, apparently related to the frequency of the stimulus items, that may have effectively masked differences between interruptions that occurred within the onset and those that occurred between the onset and the rime and that may have also conceivably masked an interaction of the sonority hierarchy with syllable structure. In any event, given the pattern of our other results, we would predict an onset-rime boundary effect to emerge only postlexically, in a lexical decision task or other task requiring maintenance of information in short-term memory.

Fowler (1987) and Browman and Goldstein (1988) have argued that the syllable's internal structure may arise as a result of articulatory constraints on the timing of initial versus final consonants with respect to vowels in the same syllable. Because the phonological code required to maintain information in short-term memory is more strongly affected by articulatory suppression than the phonological code permitting lexical access (according to Besner and Davelaar, 1982), it would seem reasonable to suggest that it too has an articulatory basis (see, e.g., Hintzman, 1967).

In any event, the results of our experiments taken in conjunction with prior psycholinguistic research on the internal structure of the syllable and the sonority hierarchy would suggest the following: Support for the hierarchical structure of the syllable and for the influence of the sonority hierarchy on such structure is most likely to emerge in tasks that implicate phonological coding in short-term memory.

REFERENCES

Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology, 36*, 701-711.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140-155.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychopnysics. 39*, 187-196.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, 16, 231-241.

Davis, S. (1987). On the arguments for syllable-internal structure. Research on Speech Perception, Progress Report No. 13, Indiana University.

Derwing, B., Nearey, T. M., & Dow, M. L. (1987). On the structure of the vowel nucleus: Experimental evidence. Paper presented at the LSA Conference, San Francisco.

Dow, M. L. (1987). *The psychological reality of sub-syllabic units.* Doctoral dissertation, University of Alberta.

Fowler, C. A. (1987). Consonant-vowel cohesiveness in speech production as revealed by initial and final consonant exchanges. *Speech Communication, 6*, 231-244.

Hooper, J. B. (1976). *An introduction to natural generative phonology.* New York: Academic Press.

Hintzman, D. L. (1967). Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6*, 312-367.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements, 1*, 13-32.

Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. (1965). Tables of trigram frequency counts for various word-length and 'etter-position combinations. *Psychonomic Monograph Supplements, 1*, 33-78.

MacKay, D. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.

McCusker, L., Hillinger, M., & Bias, R. (1981). Phonological recoding and reading. *Psychological Bulletin, 89*, 217-245.

McNamara, T., & Healy, A. (1988). Semantic, phonological, and mediated priming in reading and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 398-409.

Pechmann, T., Reetz, H., & Zerbst, D. (1989) Kritik enier Messmethode: Zur Ungenauigkeit von voice-key Messungen. *Sprache & Kongition, 8*, 65-71.

Seidenberg, M. S., (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, 12, 315-323.

Selkirk, E. O. (1982). The syllable. In H. Van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Part II). Dordecht, Holland: Foris.

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review.* Bloomington: Indiana University Linguistics Club.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*, 49-74.

Treiman, R. (1984). On the status of final consonant clusters in English syllables. *Journal of Verbal Learning and Verbal Behavior, 23*, 343-356.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25*, 476-491.

Treiman, R. & Chafetz, J. (1987). Are there onset- and rime- like units in printed words? In M. Coltheart (Ed.), *Attention and performance XXII.* London: Erlbaum.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 145-152.

Van Orden, G. (in press). Phonological mediation is fundamental to reading. In D. Besner & G. Humphreys, (Eds.), *Basic processes in reading: Visual word recognition.* Hillsdale, NJ: Erlbaum.

Wike, E. L., & Church, J. D. (1976). Comments on Clark's "The Language-as-fixed-effect fallacy." *Journal of Verbal Learning and Verbal Behavior, 15*, 249-255.

FOOTNOTES

*Appears in *Journal of Psycholinguistic Research, Vol. 20(4)*, 337-363 (1991).

†Also French Department, Wellesley College.

††University of Colorado, Boulder.

†††University of Colorado. Now at Widener University, Chester, PA.

[1]See Selkirk (1982) for theoretical arguments that the syllable is hierarchically organized, but see Davis (1987) for arguments that the syllable is divided, nonhierarchically, into onset, peak and coda.

[2]Four of the six nasal items and one of the six obstruent items had a $C_1C_2VC_3$ structure, although all were five letters long.

[3]Three of the items in this group began with the letters "ch," characterized by some phonologists as a single phoneme /č/ and by others as a sequence of two phonemes /tʃ/.

[4]We inadvertently included two items that appeared both as onset-rime and as nucleus-coda test items, stern and brand. The associated nonwords were different in each case.

[5]We do not report the results of our analysis of the nonword data because the significant effects provided no support for syllable-internal structure or sonority and were inconsistent across the two experiments. A comparison of the speed analysis and the error analysis also indicated a number of probable speed-accuracy tradeoffs, although these were not evident in the word data.

[6]This transformation produced more normally distributed values and eliminated disproportionate influences by outliers.

[7]Although there were differences in frequency in the onset-rime groups, these differences were not significant, $F(3,32)=.162$, $p = .9208$, MSe=69.8620. Nonetheless, we believe that the onset effect is best explained in terms of word frequency. We examined single-letter and di- and trigram frequencies (Mayzner & Tresselt, 1965; Mayzner, Tresselt, &

Wolin, 1965) and found no correlation with the pattern of our results for onset-rime (or coda) test words. Furthermore, both the word and the nonword stimuli had the same initial consonant clusters, but the onset effect only occurred in the word data. Finally, as suggested by an anonymous reviewer, we compared the mean latencies of the subjects in our two experiments to s+n onset-rime words (which are relatively infrequent) and s+m onset-time words (which are relatively frequent) and found a significant frequency effect there as well, $t(29)=3.188$, $p=.0034$, two tailed.

APPENDIX

Test Items Used in Experiments 1 and 2

Onset-Rime

Word	Nonword	Word	Nonword	Word	Nonword
c_1c_2=1 phoneme		c_2=liquid		c_2=nasal	
chest	chorn	craft	flaft	smart	snart
thorn	thest	flint	crint	smash	snash
chill	chigh	drank	glank	sniff	smiff
thigh	thill	glint	drint	snarl	smarl
shark	sheft	clasp	blasp	smell	snell
theft	thark	blend	clend	snuff	smuff
shunt	shump	prank	trank	c_2=obstruent	
chump	chunt	tramp	pramp	stern	spern
shawl	chawl	plump	brump	spasm	stasm
champ	shamp	brand	pland	skunk	scunk
thumb	shumb	clink	grink	scowl	skowl
shirt	thirt	grind	clind	stark	scark
				skimp	stimp

Nucleus-Coda

Word	Nonword	Word	Nonword	Word	Nonword
c_3=obstruent		c_3=liquid		c_3=nasal	
blast	crasp	dwarf	smarf	blunt	swunt
crisp	blisp	smirk	dwirk	swamp	blamp
brisk	crisk	scald	scort	blank	slank
crust	brust	snort	snald	slump	blump
cleft	greft	scalp	scern	print	clint
grist	clist	stern	stalp	clump	prump
draft	twaft	spark	skark	stint	blint
twist	drist	skirt	spirt	blond	stond
tract	traft	sport	sporm	blink	blant
graft	gract	storm	stort	scant	scink
grasp	frasp	spurt	spirl	trunk	trand
frost	grost	swirl	swurt	brand	brunk

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 89-108

Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants*

Catherine T. Best† and Winifred Strange††

Past research suggests that the degree of difficulty adults have with discriminating nonnative segmental contrasts varies considerably across contrasts and languages. According to a recent proposal, this variation may be explained by differences in how the nonnative phones are perceptually assimilated into native phoneme categories (Best, McRoberts & Sithole, 1988). The present study examined that proposal by testing identification and discrimination of three synthetic series of American English approximant contrasts, presented to American English-speaking subjects and native Japanese-speaking learners of English. The English approximants differ with respect to their phonemic status in Japanese, as well as in the phonetic details of the most similar Japanese phonemes. The perceptual assimilation hypotheses were strongly upheld in cross-language comparisons. Moreover, on the assumption that perceptual assimilation may be modified by learning the second language (L2), we also evaluated differences between subgroups of the Japanese subjects who had two different levels of English conversation experience. Those with intensive English conversation experience showed identification and discrimination patterns that were more similar (but not identical) to the Americans' performance than did those who had had little English experience.

1. INTRODUCTION

Language-specific experience influences the perception of phoneme contrasts. Adults are often hampered in their identification and/or discrimination of phones that are not employed contrastively in the phonological system of their language. For example, monolingual Japanese and Korean speakers have difficulty distinguishing the American English liquids /r/ and /l/, which do not occur contrastively in their native languages (Gillette, 1980; Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975; Sheldon & Strange, 1982). Analogously, English speakers have difficulty with some nonnative contrasts such as the Czech retroflex vs. palatal fricatives (Trehub, 1976), Thai voiced vs. voiceless unaspirated stops (Lisker & Abramson, 1970), Hindi dental vs. retroflex stops, and Salish velar vs. uvular ejectives (Polka, 1991; Tees & Werker, 1984; Werker & Tees, 1984). This perceptual difficulty, however, appears to be neither universal nor immutable. Some nonnative contrasts are relatively easy to discriminate even without prior exposure or training (e.g., Best, 1992; Best, McRoberts, & Sithole, 1988). Perceptual difficulties with particular contrasts also vary depending on syllable position and phonetic context (e.g., Mochizuki, 1981). Other contrasts are distinguishable when listening conditions minimize memory demands or phonemic categorization (Carney, Widin, & Viemeister, 1977; Werker & Logan, 1985).

This work was supported by NIH grants NS-05877 and DC-00403 to the first author and grant HD-01994 to Haskins Laboratories, as well as DC-00323 to Winifred Strange, which supported her during the preparation of this manuscript, and NIMH grant MH-21153 to James Jenkins, which supported the second author while she was a visiting scholar at Haskins Laboratories. We gratefully acknowledge Leon Seraphim for his help in establishing contact with our Japanese subjects; Arlene Antilla, Susan Koski, and Marshall Gladstone for their help with data collection; and James Jenkins for help with final data analyses. We also thank Hajime Hirose and Arthur Abramson for helpful discussions about the phonetic properties of Japanese approximants, as well as James Flege, Virginia Mann, and an anonymous reviewer for comments on an earlier draft of this manuscript.

Discrimination of nonnative contrasts that are initially difficult for adults can sometimes be improved rapidly through laboratory training (e.g., Pisoni, Aslin, Perey, & Hennessy, 1982), while others are resistant to change (Strange & Dittmann, 1984). Perception of non-native contrasts improves in the course of learning to speak a second language (L2), even in adulthood (e.g., MacKain, Best, & Strange, 1981), although improvement is often more marked if exposure to L2 occurs before puberty (Tees & Werker, 1984; Yamada & Tokhura, 1991; see Flege, 1988). Furthermore, some individuals appear to be more sensitive than others to nonnative distinctions even without experience or training (e.g., subject M. K. in MacKain et al., 1981; see also Polka, 1991; Pruitt, Strange, Polka, & Aguilar, 1990).

The fact that native language experience constrains perception of nonnative contrasts, but that further experience with nonnative sounds may nonetheless alter those perceptual constraints even in adults, raises questions about the nature of the native-language influence. Specifically, what properties do listeners perceive in nonnative sounds, and how might those properties relate to the perceived properties of native phonemes?

Recently, it has been proposed that mature listeners perceptually assimilate most nonnative phones to native categories (Best, 1992; Best et al., 1988; cf. Flege, 1990). That is, the nonnative phones are perceived in terms of their similarities (and dissimilarities) to native phonemes. According to this model, mature language users assimilate nonnative speech sounds to native categories on the basis of their perceived gestural (articulatory-phonetic) similarities to native phones (Best, 1992). The gestural similarities and dissimilarities referred to are based on the model of gestural phonology proposed by Browman and Goldstein (e.g., 1986, 1989; Goldstein & Browman, 1986), i.e., they refer to temporal and spatial properties (i.e., degree and location of constrictions) of the dynamic movements of vocal tract articulators such as lips, jaw, tongue body, glottis, etc.

Four perceptual assimilation patterns are possible: 1) The two members of the nonnative contrast may be assimilated into **two categories** in the native phonology; 2) Both nonnative phones may be assimilated equally well (or poorly) into a **single category**; 3) Both may be assimilated into a single category, but unequally, thus showing a **category goodness difference** in their fit to the native phoneme; or 4) The nonnative phones may differ so much from the phonetic properties of native phonemes that they are **non-assimilable**. Note that the assimilation pattern depends on the listener's *perception* of similarities; listeners may differ from one another, even within the same native language, with respect to which phonetic properties of a nonnative phone they may detect or attend to in perception. (Although it might be argued that nonnative phones are assimilated on the basis of *acoustic*-phonetic similarities rather than, or in addition to, gestural similarities, the distinction is difficult to make because articulatory- and acoustic-phonetic properties are confounded in the signal.)

Best and colleagues (1988, 1992) predicted that phones that are assimilated equally to a single category should prove most difficult to discriminate. Discrimination of phones assimilated to two different native categories should be quite good, while contrasts that are non-assimilable, or those that show a category goodness difference in assimilation, should result in intermediate and variable levels of discrimination difficulty. The level of discrimination for nonnative phones that differ in category goodness should depend on the degree of perceived *phonetic* similarity between the native phoneme category and each of the nonnative phone categories. Non-assimilable contrasts are perceived as nonspeech sounds rather than as phonological segments; for them, discrimination difficulty should be a function of *acoustic* similarity.

Thus, the issue of native-language (L1) influence on perception of nonnative speech contrasts focuses on the relation between phonetic details and phonemic categories. In turn, any readjustment in perception as a result of further experience with nonnative phones would seem to involve an adjustment in the perceived phonetic details of the second language (L2) phoneme categories (cf., Flege, 1990; Flege & Bohn, 1989). That is, nonnative phones may be assimilated to native phonemes to the strongest degree by listeners who have had little or no L2 experience. However, increased L2 experience may foster improved recognition of the *discrepancies* between the L1 and L2 phones. This could lead to a decline in degree of assimilation of L2 phones to L1 categories, and perhaps ultimately to the emergence of a separate L2 phoneme category due to improved recognition of phonetic properties within the L2 phonological system. We pursued these issues in the present study by examining the perception of three English approximant contrasts by American English listeners and by Japanese listeners at two levels of English experience.

Contrasts between approximant consonants (/w-j/, /w-r/, and /r-l/) in syllable-initial position offer a rich context for studying the perceptual influence of both phonetic and phonological differences between American English and Japanese. The contrasts differ across these languages in their phonological status; /r-l/ is a phonemic contrast in English but not in Japanese. The remaining two contrasts can be said to represent abstract phonological oppositions in both languages. However, /w-j/ and /w-r/ differ in terms of the similarities between American and Japanese phonetic realizations of the phonemic categories.

Realizations of /j/ are quite similar in the two languages, differing only slightly in phonetic and phonotactic details. Both are glide consonants with a palatal place of articulation and spread or neutral lip posture. However, Japanese phonotactic constraints disallow the occurrence of /j/ before the high front vowels /i/ and /e/, whereas no such restrictions occur in English. Also, the starting tongue posture has been described as somewhat lower and further back for Japanese /j/ (Vance, 1987) than for English /j/ preceding /a/ (the context used in this study), which should, if true, result in slightly higher F1 and lower F2 and F3 onset frequencies for Japanese /j/.

The phonetic realization of /w/ differs more obviously between languages. In English, /w/ is realized with lip-rounding or protrusion ([w]), similar to the back rounded English vowel /u/, whereas in Japanese, /w/ is produced with spread lips ([ɰ]), similar to the back unrounded Japanese vowel [ɯ] (Bloch, 1950; Vance, 1987). Because lip rounding/protrusion lowers the frequency of all formants (especially upper formants), F2 and F3 onset frequencies should be higher (hence more similar to English /j/) in Japanese than English (see Kasuya, Takeuchi, Sato, & Kido, 1982; Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957).

The cross-language discrepancy in the phonetic realization of /r/ is even greater, involving a difference in both manner of articulation and tongue posture. Whereas American English /r/ is a retroflex or palato-alveolar central approximant ([ɻ] or [ɹ], respectively), Japanese /r/ is usually an alveolar tap [ɾ] rather than an approximant. (Bloch, 1950; Price, 1981; Vance, 1987). In addition, while English /l/ is an alveolar lateral approximant, Japanese does not employ a distinct /l/ phoneme. Japanese /r/ is, in fact, variably pronounced, and is occasionally realized in some positions by some speakers as an approximant [ɻ] or [ɹ], as a retroflex stop [ɖ], as an alveolar trill [r], or even as a lateral alveolar tap [ɺ]. Thus, the lateral alveolar is a rare allophone of /r/ in Japanese and is apparently not even then an approximant; rhotic approximants may occur but are also quite rare (Bloch, 1950; Miyawaki, 1973; Vance, 1987).

According to the perceptual assimilation model (Best et al., 1988; 1992), Japanese listeners would be expected to assimilate the English /w-j/ contrast as a two category contrast *vis a vis* their native phonology. However, the phonetic boundary between categories may be shifted toward /j/ (that is, Japanese may hear more /w/s), since the Japanese /w/ is unrounded and is more similar to English /j/ acoustically and articulatorily than is the American English /w/. Nonetheless, categorization and discrimination should be quite good. English /w-r/ might be expected to be assimilated to a single Japanese phoneme category, but as a contrast involving a category goodness difference. That is, since English /r/ is an approximant, not a tap as in Japanese, it seems likely to be assimilated as a "poor" exemplar of the Japanese approximant /w/, whereas English /w/ would be assimilated as a "better" exemplar of Japanese /w/. The possibility that [ɹ] would assimilate to Japanese /w/ is supported by evidence from Mochizuki (1981) and Yamada and Tokhura (1991). The alternative possibility, though less likely, is that English /r/ might be assimilated as a very poor exemplar of the Japanese tapped /r /, which would lead to two category assimilation for /w-r/. In either case, Japanese discrimination of /w-r/ should be good. Finally, English /r-l/ should result in single category assimilation by Japanese, in which both phones are equivalently poor exemplars either of their approximant /w/ or (less likely) of their tapped /r/. Japanese categorization and discrimination are known to be rather poor for syllable-initial /r/ and /l/, particularly for those who have had little conversational English experience (Miyawaki et al., 1975; Mochizuki, 1981).

Best et al. (1988; 1992) discussed assimilation of nonnative speech contrasts only in terms of their relative levels of discriminability. In the present study, the concept of perceptual assimilation was extended to predict cross-language differences in phonetic category boundaries along synthetic approximant series that interpolated on multiple, phonetically-relevant acoustic parameters. Specifically, in identification tests of /w-j/ and /w-r/ series, the Japanese listeners were expected to label more of the acoustically intermediate stimuli as /w/ than American listeners. For /w-j/, which are distinguished primarily by F2 and F3 onsets

and transitions, stimuli with higher F2 and F3 values are more similar to Japanese [ɰ] than to American [w]. Thus, the Japanese /w-j/ boundary should be shifted toward /j/, relative to the American boundary. However, the steepness of the category boundary should be equivalent in the two language groups because the contrast reflects a phonological opposition for both.

In the case of /w-r/, Japanese listeners might be expected to label more intermediate stimuli as /w/ rather than as /r/, as compared to American listeners, because the slow transitions of these approximants are more similar to the Japanese /w/ than to their tapped /r/ (see also Mochizuki, 1981). Yet because neither the English /w/ nor /r/ are ideal exemplars of Japanese phoneme categories, and because /w-r/ was expected to be assimilated as a category goodness difference within the Japanese /w/ category, their identification function was expected to be less steep in the region of the category boundary than that of American listeners.

No clear predictions can be made about the *location* of the /r-l/ boundary for Japanese. However, the predicted single category assimilation pattern is consistent with previous findings that the labeling function is less clearly-defined for Japanese than for American listeners, resulting in a shallower slope at the category boundary (e.g., MacKain et al., 1981; Miyawaki et al., 1975).

If increased L2 experience serves to shift adults' perception of the phonetic details of nonnative phonemes toward improved recognition of the discrepancies between L2 phones and the L1 categories to which they were initially assimilated (cf. Flege, 1989; 1990), additional predictions can be made about relative performance on the three contrasts by Japanese subjects with more or less spoken English experience. According to perceptual assimilation predictions (Best et al., 1988; 1992), Japanese listeners with little English experience should discriminate the /w-j/ contrast best, as a two category contrast, with a peak in discrimination functions at their category boundary (i.e., shifted toward the /j/ end of the series). They should show lower discrimination levels and a lower, broader boundary-related peak (also shifted toward /r/) in discrimination of the English /w-r/ contrast, which shows a category goodness difference with respect to Japanese /w/. Their discrimination should be poorer still on the English /r-l/ contrast, a single category assimilation type. Thus, discrimination performance by inexperienced Japanese listeners should be equivalent to that of American listeners on the /w-j/ contrast, somewhat lower on the /w-r/ contrast, and perhaps even lower on the /r-l/ contrast. In comparing identification performance of Americans and the two Japanese subgroups, we expected that category boundary steepness for /w-j/ would be equivalent across all three groups, but less steep for the inexperienced Japanese than the other two groups on the /w-r/ and /r-l/ series. Japanese with more extensive English conversational training were expected to discriminate and identify all three contrasts in a pattern more similar to that of American adults than their peers who had had minimal English experience, i.e., the position and steepness of their category boundaries should have become shifted toward the values found in Americans. However, according to earlier work showing residual differences from Americans on syllable-initial /r-l/ (MacKain et al., 1981), even the experienced Japanese listeners were expected to differ somewhat from the Americans on the /w-r/ and /r-l/ series in both boundary position and steepness, as well as in discrimination levels.

2. EXPERIMENT 1

2.1 Method

The aim of this study was to compare identification and discrimination of synthetic /r-l/, /w-r/, and /w-j/ series by American and Japanese listeners. A previous report had examined perception of an /r-l/ series by these two language groups (MacKain et al., 1981). The stimuli and methods for the /r-l/ tasks, as well as the results for a larger group of Japanese subjects on that contrast, were presented in the earlier publication. For the present paper, we reanalyzed a subset of those earlier-reported data for comparison with responses of the same listeners on the other two approximant series.

2.1.1 Subjects. Nine of the 10 original American participants in the MacKain et al. study returned within the subsequent two weeks for two additional test sessions on the /w-r/ and /w-j/ contrasts. All were college undergraduates (4 males, 5 females) recruited through notices posted at Yale University.

Nine of the 13 Japanese who participated in the original study returned within two weeks for tests on the other two approximant contrasts. Four Japanese (2 males, 2 females) had had intensive English conversational instruction with native American English speakers (8-10 hours/week) and had been in residence in the USA for 18 to 48 months at the time of testing (Ss 7-10 in MacKain,

et al., 1981). These subjects are hereafter referred to as the Experienced Japanese. Five others (4 male, 1 female) had had little or no English conversational instruction (0-3 hours/week) and had resided in the USA less than 7 months (Ss1-4 and S13 in MacKain, et al., 1981). These are hereafter referred to as the Inexperienced Japanese. Note that S13 was subject M. K., an anomalous listener who showed remarkably good /r-l/ perception even though he had been in the U. S. only briefly and had had little conversational experience with English. He was discussed separately in MacKain et al., but was incorporated into the Inexperienced group for the present study because of the small number of subjects in each subgroup.

All subjects were paid. All reported good hearing in both ears and could read written English.

2.1.2 Stimulus Materials. The /r-l/ series was a /rak/-/lak/ continuum, and is described in detail in MacKain et al. (1981). Two additional series, /wak/-/jak/ and /wak/-/rak/, were generated in analogous manner on the OVE-IIIc cascade formant synthesizer at Haskins Laboratories. Synthesis parameters for series endpoints, /jak/, /wak/, /rak/ (and /lak/), were derived from an analysis of real speech tokens produced by an adult male speaker of American English. These endpoint synthetic stimuli were equated for overall duration (330 ms including the silence and burst of a natural /k/), amplitude and intonation contour (rising-falling), and spectral pattern of the final 105 ms of the 210 ms vocalic portion of the syllable. The initial 105 ms of the four stimuli differed in frequency of onset and the subsequent pattern of transitions of the first three oral formants (F1, F2, F3, respectively). Table 1 gives the onset frequencies of these formants for the four endpoint stimuli, and Figure 1 provides a schematic diagram of the formant patterns for the endpoint stimuli of each continuum.

Table 1. *Nominal stimulus parameters for endpoint stimuli.*

	Formant Onset Frequencies (Hz)		
Stimuli	F1	F2	F3
/jak/	275	2105	2809
		*[a]	*
/wak/	275	644	2295
	*	*	*
/rak/	349	1067	1477
		*	*
/lak/	349	1207	?594

[a]Asterisk indicates that the parameters are interpolated to produce series between endpoints.

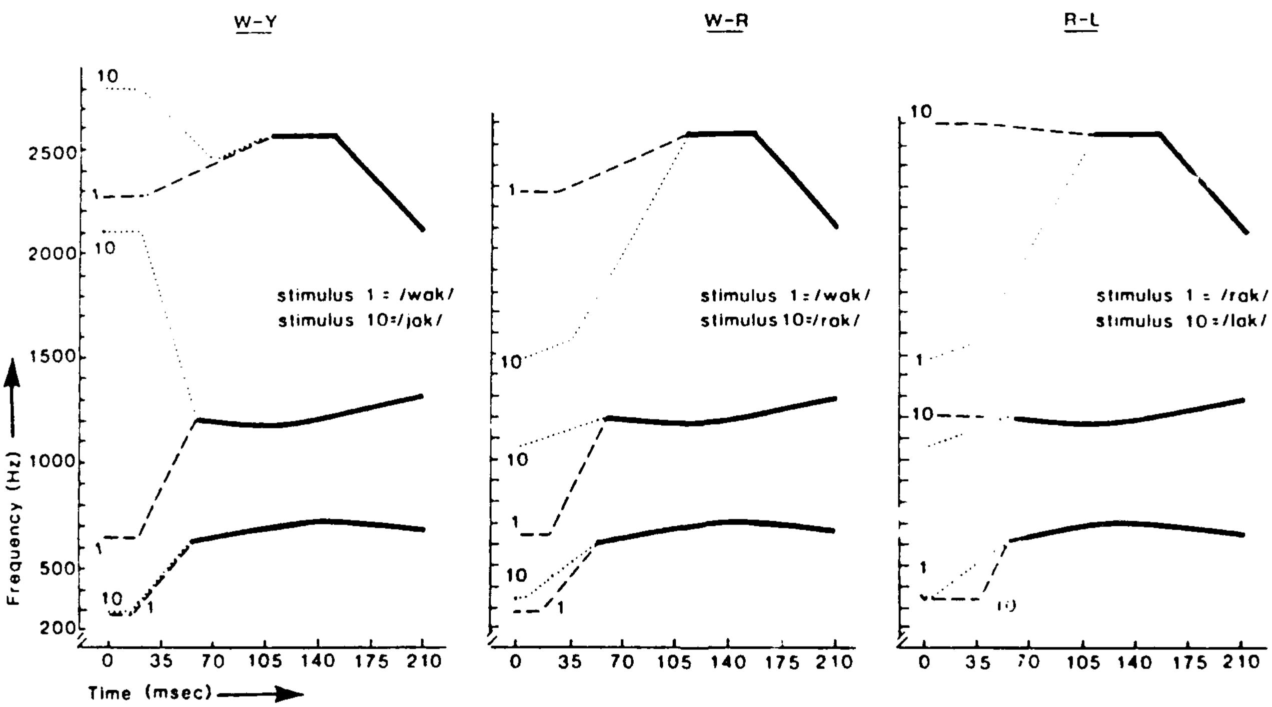

Figure 1. **Schematic diagram of the center frequencies of F1, F2, and F3 in the endpoint stimuli for the three stimulus series.**

The 10-step /wak/-/jak/ series was generated by interpolating on the F2 and F3 onset frequencies in approximately equal steps of 162 Hz and 57 Hz, respectively, from the /wak/ pattern (item 1) to the /jak/ pattern (item 10). The initial steady-state portion was 28 ms for F2. F3 was steady-state for 21 ms, followed by a linear transition of 49 ms to a common frequency (2379 Hz). As can be seen in Figure 1, this produced a "dip" in F3 for stimuli toward the /jak/ end of the series, which is characteristic of /j/ in natural utterances.

The 10-step /wak/-/rak/ series was generated by interpolating between /wak/ (item 1) and /rak/ (item 10) on F1, F2, and F3 onset frequency (and subsequent transitions) in approximately equal steps of 8 Hz, 47 Hz, and 91 Hz, respectively. An inflection point 28 ms after onset of F2 and F3, and 21 ms after onset for F1, produced an initial quasi-steady-state pattern (see Figure 1).

For comparison, the endpoints of the /rak/-/lak/ series are included in Table 1 and in Figure 1. In this series, onsets and transitions of F2 and F3 were varied, as well as the temporal pattern of the F1 transition (See MacKain et al., 1981, for a detailed description).

2.1.3 Procedure. The tests for the /rak/-lak/ series are described in MacKain et al. (1981). The tests for the other two series were similar in format, except that the oddity discrimination test used in the previous study was not employed; only the AXB discrimination task was used for the present report. All subjects completed two sessions consisting of two tests each, with a 15-minute break between the first and second test of the session. In one session subjects completed a 2-choice forced choice identification test followed by an AXB discrimination test of the /w-j/ series. The other session included identification and AXB discrimination tests of the /w-r/ series. Testing was conducted in a sound-attenuated chamber with 2-4 subjects at a time (all from a single language group during a given test session). Subjects listened over headphones (Telephonics TDH-39) to stimuli presented via a Crown reel-to-reel tape deck at a comfortable loudness level (approximately 75 dB SPL).

Each identification test included 20 repetitions of each of the 10 stimuli in the series being tested, presented singly and randomized within each block of 10 trials. Intertrial intervals (ITIs) were 2.5 s; interblock intervals (IBIs) were 4 s. For each trial, subjects were asked to write one of two letters to indicate the initial consonant of the syllables they heard; that is they wrote "W" or "Y" during the /w-j/ identification tests, and "W" or "R" during the /w-r/ identification tests.

The AXB discrimination procedure was chosen because of its relatively low memory demands and low sensitivity to observer bias, by comparison to other standard discrimination procedures such as oddity, 2IAX and 4IAX (e.g., Best, Morrongiello, & Robson, 1981; MacKain et al, 1981; cf. Pollack & Pisoni, 1971). Each AXB discrimination test contained 10 repetitions of each of the 2 AXB orders for the 7 possible pairings of stimuli that differed by 3 steps along the continuum being tested (1-4, 2-5, 3-6, 4-7, 5-8, 6-9, and 7-10). Trials occurred in blocks of 14 (2 orders × 7 AXB pairings), and were randomized within blocks. Within-trial interstimulus intervals (ISIs) were 1 s, ITIs were 3 s, and IBIs were 6 s. For each trial, the subject circled the number "1" or the number "3" to indicate whether the second item of the trial (X) matched the first (A) or the third (B) item of that trial.

2.2 Results. The results of identification tests are reported first, followed by the results of discrimination tests. Differences between the American group and the Japanese group as a whole were statistically analyzed. Performance by Experienced and Inexperienced Japanese subgroups were compared with the American group in separate analyses. For all analyses, data on the perception of /r-l/ by the 9 Americans and 9 Japanese, which were a subset of the data reported previously in MacKain et al. (1981), were included for comparison with results on the /w-r/ and /w-j/ series.

2.2.1 Identification tests. Figure 2 presents the pooled identification functions for the American and Japanese groups on the /w-j/, the /w-r/, and the /r-l/ continua. These functions represent the raw identification data, averaged over 9 subjects in each group. As the figure shows, the American listeners labeled /w-j/ and /w-r/ categorically, with abrupt crossovers at category boundaries and highly consistent labeling of within-category stimuli. Performance was commensurate with their identification of the /r-l/ series. The Japanese as a group also labeled /w-j/ and /w-r/ categorically. This contrasts with their identification performance on the /r-l/ series, which showed less consistency in labeling within-category stimuli. As previously reported, performance by the Japanese was markedly different from that of the American listeners on the /r-l/ series.

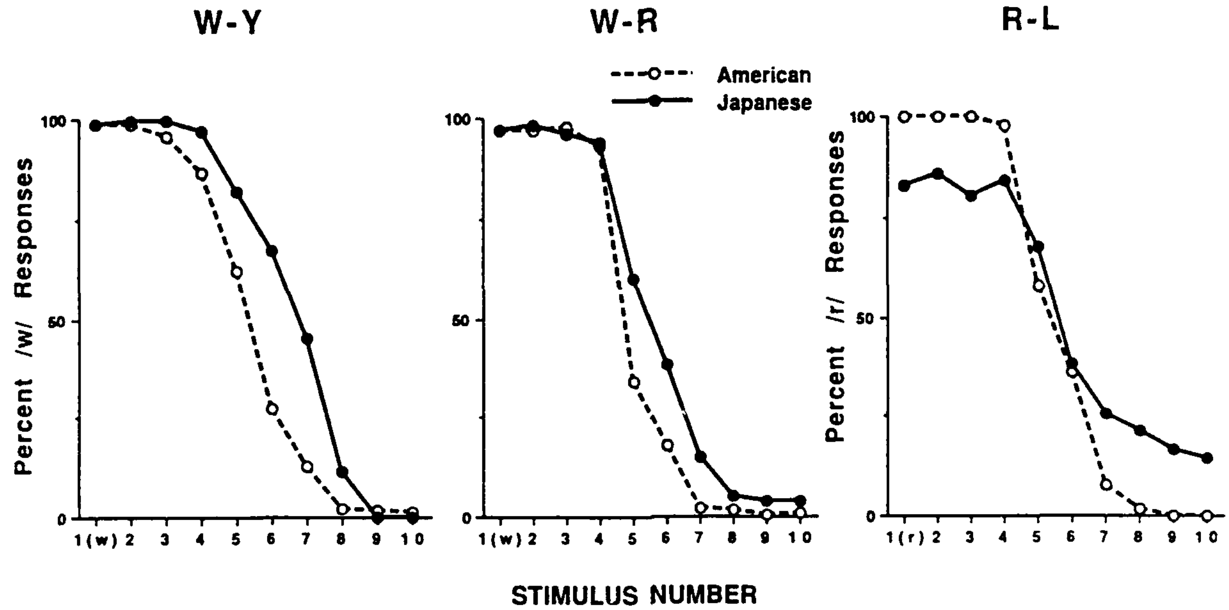

Figure 2. **Average identification functions for the American and Japanese listener groups on the three series.**

In order to make between-group comparisons on the location and steepness of category boundaries for the three series, best fit ogives of individual subjects' identification functions were determined through narrow-range PROBIT analyses, using the labeling probabilities on the three stimuli closest to the 50% crossover. This statistical procedure fits a cumulative normal curve to the raw data, thus smoothing the function. Category boundaries were defined as the 50% intercept of the ogives. The slopes of these ogives (1/s.d.) indicate the peak rate of change in category labeling at the crossover, and were used as a reflection of the steepness of the category boundaries, i.e., larger slope values indicate steeper functions.

The ogives for the Americans and the two Japanese subgroups are displayed in Figure 3. χ^2 values were significant, indicating a significant deviation between the raw data and the fitted ogives, for only 6 out of the 54 PROBIT analyses (2 groups x 9 subjects x 3 series): three Americans on /w-r/, one American on /r-l/, and two Experienced Japanese on /w-j/. In all cases, the significant χ^2 resulted from extremely sharp category boundaries that were not well-fitted to three data points, and would have fit better for two points. There were only two cases of grossly nonmonotonic raw identification functions for two Inexperienced Japanese on /r-l/. In neither case was the PROBIT χ^2 significant, i.e., the ogives provided a good fit to the raw data.

2.2.1.1 Boundary location analyses. The boundary locations for American and Japanese groups (expressed in terms of stimulus number) on each series are given in Table 2. These data indicate that, on average, the boundaries for the Japanese on all three series fell to the right of the American boundaries. That is, the mean boundary values show that the Japanese labeled more stimuli as /w/ on the /w-j/ and /w-r/ series, and more stimuli as /r/ on the /r-l/ series. Note also that the variability of boundary locations appears to be greater on the /w-j/ series than on the /w-r/ series for both Japanese and American subjects, as reflected in the standard deviations (SD's).

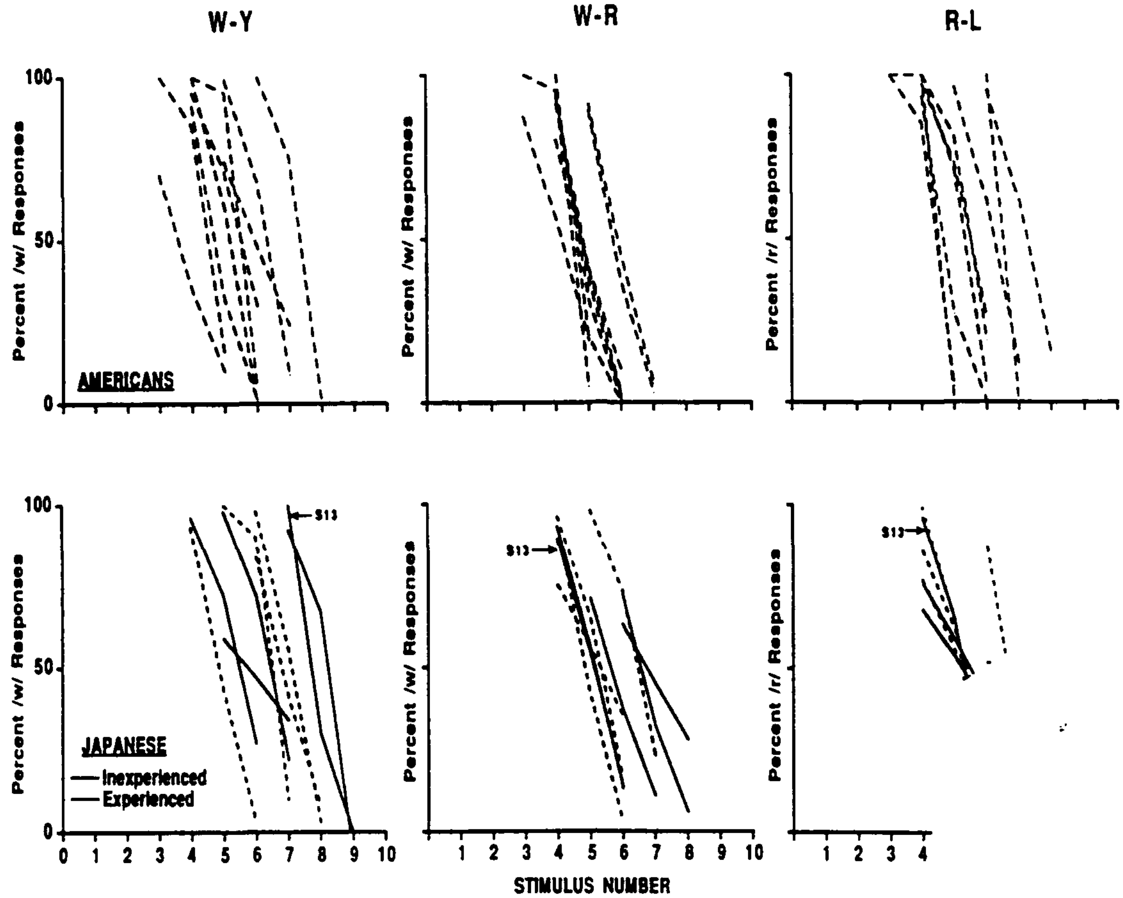

Figure 3. **Narrow-range fitted ogive functions for individual subjects in the American and Japanese groups. The S13 lines indicated in the Japanese plots refer to the data from subject M. K., discussed in MacKain et al. (1981) as being inexperienced with American English conversation yet similar to Americans in categorization of /r/ and /l/.**

To test the reliability of these boundary differences, a Groups (American vs. Japanese) × Series (/w-j/, /w-r/, /r-l/) analysis of variance (ANOVA) of the 50% intercept values of best fit ogives for individual subjects was conducted. The main effect of Groups was significant, $F(1,16) = 10.82$, $p < .005$, indicating that the Japanese boundaries were indeed shifted significantly rightward in comparison to the American boundaries. Neither the Series main effect nor the Groups × Series interaction approached significance (p's = .17 and .64, respectively), suggesting that the rightward shift of the Japanese boundary occurred in all three series, and to approximately the same degree in each. However, *a priori* predictions about possible cross-language differences on the boundaries for each series warranted an analysis of simple effects, which indicated that the language difference was significant for /w-j/, $F(1,48) = 6.44$, $p < .02$, but was marginal for /w-r/($p = .10$) and nonsignificant for /r-l/ ($p = .24$). That is, the boundary shift between language groups was reliable only for /w-j/.

To assess the statistical reliability of the differences between Experienced and Inexperienced subgroups in comparison with American listeners, an English Experience (American vs. Experienced Japanese vs. Inexperienced Japanese) × Series ANOVA was computed. (Because group sizes were small and unequal, these statistical results should be interpreted cautiously, although these factors decrease rather than increase the likelihood of attaining statistical significance.) The main effect of English Experience was significant, $F(2,15) = 6.75$, $p < .01$, while the main effect of Series and the English Experience × Series interaction were nonsignificant. Planned linear contrasts among the three groups, based on *a priori* predictions, yielded reliable evidence that the boundary for the Experienced Japanese subjects was intermediate between that of the Americans and that of the Inexperienced Japanese, $F(1,15) = 13.12$, $p < .003$. Table 2 summarizes these differences in boundary locations for the Experienced and Inexperienced Japanese subjects.

2.2.1.2 Slope analyses. Table 3 presents the data on steepness of category boundaries for American and Japanese groups (expressed as the mean slope of their ogives). The Japanese showed a pattern across the three series that was strikingly different from the Americans. The slope for /w-j/ was steepest and most similar to Americans', while those for /w-r/ and /r-l/ were less steep than Americans'. This was as predicted on the reasoning that /w-j/ would constitute a two category distinction for the Japanese, while /w-r/ would show a category goodness difference within a Japanese category, and both /r/ and /l/ would show a poor fit to one Japanese category.

The statistical reliability of these differences was assessed in a Groups (American vs. Japanese) × Series ANOVA of slope values. The main effect of Groups was significant, $F(1,16) = 5.47$, $p < .04$, indicating that, overall, the American boundaries were significantly more abrupt than the Japanese boundaries. Neither the Series main effect nor the Groups × Series interaction was significant. However, *a priori* predictions about cross-language differences warranted simple effects tests, which indicated that the American slopes were steeper than the Japanese slopes on /w-r/, $F(1,16) = 5.77$, $p < .03$, and /r-l/, $F(1,16) = 11.58$, $p < .04$, but not on /w-j/ ($p = .80$).

Again, the Japanese data for Experienced and Inexperienced subjects were analysed in an English Experience × Series ANOVA which included comparisons to the American group. Although the main effect of English Experience was only marginally significant, $F(2,15) = 2.91$, $p < .09$, planned linear contrasts were warranted by *a priori* predictions (American > Experienced Japanese > Inexperienced Japanese). These tests revealed the predicted direction of group differences was significant for /r-l/, $F(1,15) = 7.36$, $p < .02$, and /w-r/, $F(1,15) = 5.03$, $p < .05$, but not for /w-j/ ($p = .99$), all as expected. No other effects were significant.

To summarize, the Japanese /w-j/ boundary was shifted toward /j/ relative to the American boundary. Both Experienced and Inexperienced Japanese labeled more intermediate stimuli as /w/ than the Americans, as predicted from cross-language differences in the phonetic details of /w/. Also as predicted, the steepness of the category boundary slope on this series did not differ betw[illegible] language groups, indicating that the di[illegible] between /w/ and /j/ categories was equally s[illegible]p for all groups of listeners. These findings suggest that the American /w-j/ distinction was assimilated as a two category contrast by the Japanese listeners, with /w/-like and acoustically intermediate stimuli assimilating to the phonetically different Japanese /w/, and /j/-like stimuli assimilating to the phonetically similar Japanese /j/ phoneme category. This characterization is somewhat qualified, however, by the discrimination results on /w-j/ (see below).

Table 2. *Boundary locations for American English and Japanese listeners, including Japanese subgroups. Numerical values represent stimulus numbers along each of the test series.*

	/w-j/ mean	(SD)	/w-r/ mean	(SD)	/r-l/ mean	(SD)
Americans	5.36	(1.05)	4.93	(0.57)	5.53	(0.96)
Japanese: Overall	6.55	(1.07)	5.72	(0.70)	6.08	(1.40)
Experienced	6.32	(0.98)	5.59	(0.69)	5.60	(0.74)
Inexperienced	6.73	(1.22)	5.82	(0.77)	6.47	(1.76)

Table 3. *Slope values for American and Japanese listeners, including Japanese subgroups. Numerical values represent the peak rate of change in category responses per step along each stimulus series.*

	/w-j/ mean	(SD)	/w-r/ mean	(SD)	/r-l/ mean	(SD)
Americans	2.19	(1.29)	1.99	(1.05)	2.65	(1.67)
Japanese: Overall	2.04	(1.29)	1.09	(0.41)	1.04	(1.23)
Experienced	1.84	(0.57)	1.24	(0.49)	1.75	(1.55)
Inexperienced	2.20	(1.74)	0.97	(0.33)	0.48	(0.55)

The identification results were different for the /w-r/ and /r-l/ series than for /w-j/. As previously reported, the Japanese listeners showed significantly shallower category boundary slopes on /r/-/l/, but failed to show a significant difference in boundary location, relative to Americans. On /w-r/, the Japanese again showed a shallower boundary slope than Americans, and their boundary location differed marginally from Americans' ($p = .10$) in the predicted direction (i.e., they identified more stimuli as /w/). The /w-r/ and /r-l/ findings are consistent with the reasoning that American English /w-r/ should constitute a category-goodness difference within the Japanese /w/ category, and that English /r-l/ should represent rather poor examples of a single phoneme category in Japanese (either their glide /w/ or, less likely, their tapped /r/).

As for the effect of experience with L2, the patterns of identification performance differed as expected between the two levels of English conversation experience of the Japanese subjects. On all counts, the data of the Experienced Japanese subjects were more similar (but not identical) to the American results than were those of the Inexperienced Japanese. More intensive English conversation experience was associated with a more American-like boundary location on the English /w-j/ contrast and with steeper category boundaries for the English /w-r/ and /r-l/ contrasts.

2.2.2 Discrimination tests. Discrimination test results were also examined for evidence of native language differences and influences of L2 English experience. Percent correct responses for each of the AXB comparison pairs on each stimulus series were computed for the American and Japanese groups. Pooled discrimination functions for the Japanese and American groups are displayed in Figure 4, and mean performance levels (overall percent correct) are presented in Table 4. The relationship between American and Japanese discrimination functions varied considerably across the three series.

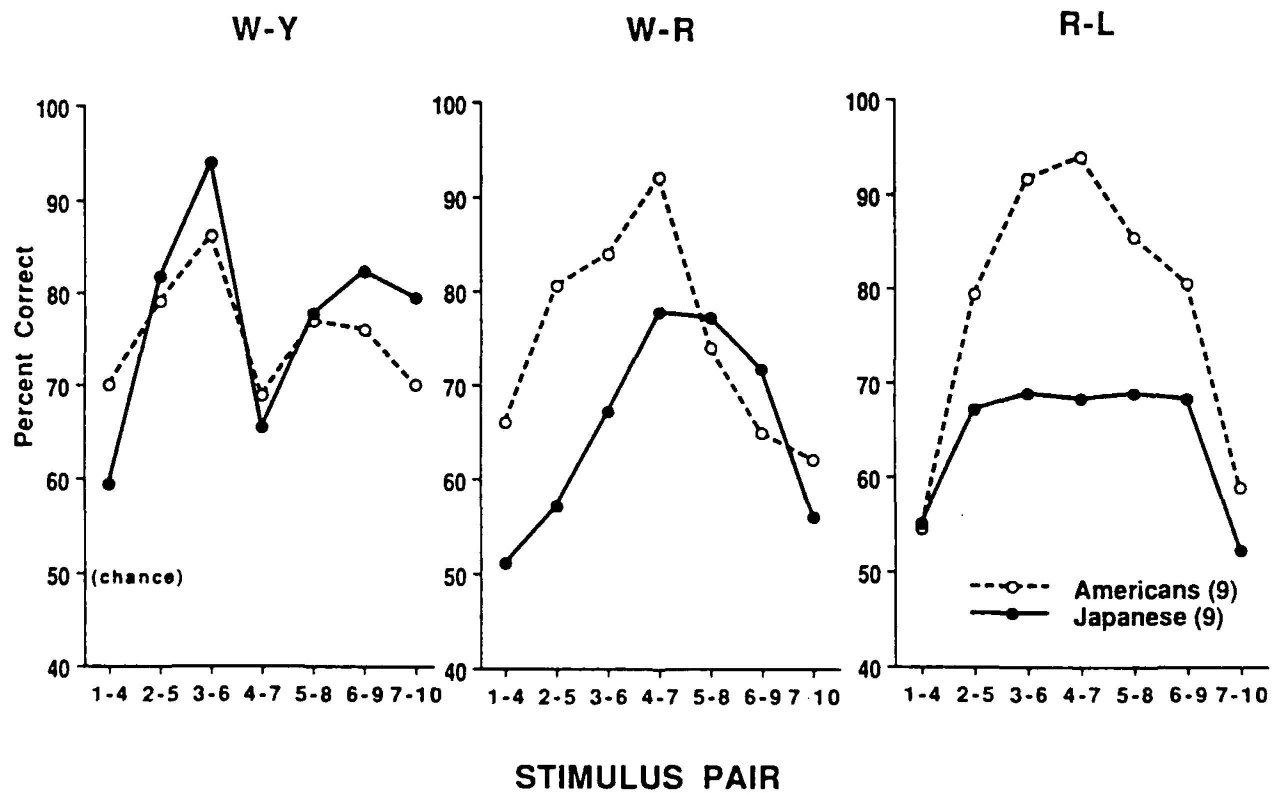

Figure 4. **Average discrimination functions for the American and Japanese groups on the three series.**

Table 4. *Mean correct performance levels pooled for American and Japanese listeners on the AXB discrimination task, including Japanese subgroups.*

	/w-j/ mean	(SD)	/w-r/ mean	(SD)	/r-l/ mean	(SD)
Americans	74.52	(11.94)	74.68	(17.46)	77.78	(19.32)
Japanese: Overall	77.14	(12.97)	65.48	(15.86)	64.13	(14.99)
Experienced	78.04	(11.57)	66.43	(18.00)	67.50	(15.55)
Inexperienced	76.43	(14.12)	64.71	(14.14)	61.43	(14.17)

The data were entered into a Groups × Series × Comparison Pairs (1-4, 2-5, 3-6, 4-7, 5-8, 6-9, 7-10) ANOVA. A significant Groups main effect, $F(1,16) = 8.55$, $p < .01$, indicated that Japanese were less accurate overall in discrimination than were Americans. The significant main effect for Comparison Pairs, $F(6,96) = 30.87$, $p < .001$, indicated that overall there were peaks and troughs in discrimination performance across the three series. The latter effect was qualified, as expected, by a Comparison Pairs × Groups interaction, $F(6,96) = 3.39$, $p < .005$, indicating that, in general, the Japanese showed smaller discrimination peaks than the American listeners. The significant Series effect, $F(2,32) = 3.64$, $p < .04$, revealed that discrimination performance was somewhat higher overall for /w-j/ than for the other two series. However, Series interacted with Group, $F(2,32) = 6.68$, $p < .004$; as expected, cross-series mean performance differed between language groups. Simple effects tests of this interaction revealed that mean performance differed among series for the Japanese, $F(2,16) = 12.77$, $p < .0005$, being substantially better for /w-j/ (77% correct) than for /w-r/ (65%) or /r-l/ (64%). Planned comparisons provided support for the order of performance that had been predicted on the basis of expected phonemic assimilation patterns (/w-j/ > /w-r/ ≥ /r-l/), $F(1,16) = 25.313$, $p < .0001$. However, a test of simple effects showed that the Americans' mean discrimination did not differ significantly across series, $p = .58$.

Comparison Pairs and Series also interacted significantly, $F(12,192) = 6.48$, $p < .001$, indicating differences in the cross-series patterns of discrimination peaks for both groups, which were further qualified by a significant Groups × Comparison Pairs × Series interaction, $F(12,192) = 3.04$, $p < 002$. To interpret these interactions, separate ANOVAs for Groups × Comparison Pairs were computed for each stimulus series. As predicted, analysis of the /w-j/ series yielded no significant difference between groups in overall discrimination accuracy. A significant main effect of Comparison Pairs, $F(6,96) = 21.14$, $p < .001$, revealed that both groups showed two peaks of relatively accurate discrimination. The occurrence of a double peak suggests that both Japanese and American listeners differentiated three rather than two categories along this synthetic continuum, although they could not indicate this in the two-category forced-choice identification test. (This possibility is considered further below and in Experiment 2.) The significant Groups × Comparison Pairs interaction, $F(6,96) = 3.46$, $p < .01$, was due to the fact that Japanese and American listeners performed differently on both within-category extremes of the series (Pairs 1-4 and 7-10). As indicated in Figure 4, Japanese subjects discriminated Pair 7-10 (within-category for /j/) more accurately, while Americans discriminated Pair 1-4 (within-category for /w/) more accurately. This asymmetry in discrimination of the endpoint within-category comparison pairs is compatible with the fact that the Japanese category boundary was shifted significantly more toward /j/ than was the American boundary. That is, both stimuli 10 and 7 fell within the /j/ category for Americans (99% and 87% of identification responses, respectively), but for the Japanese stimulus 7 was quite near the /w-j/ boundary (59% identification as /j/) while stimulus 10 was a clear /j/ (100%), which resulted in better discrimination by the latter language group. Conversely, at the other end of the series, the Japanese and Americans agreed that stimulus 1 was a clear /w/ (97 and 98%, respectively), but whereas the Japanese also identified stimulus item 4 as /w/ 98% of the time, the Americans gave only 87% /w/ identifications. Thus the Japanese

discriminated comparison pair 1-4 near chance, while the Americans discriminated that pair more readily. In fact, Americans showed the same level of performance as on pair 7-10, which had received quite similar identification scores. No other /w-j/ discrimination pairs differed between language groups.

The pattern of discrimination was quite different on the /w-r/ series. A significant Comparison Pairs effect, $F(6,96) = 9.70$, $p < .001$, reflected a single peak in discrimination performance, with troughs on either side. A significant Groups effect, $F(1,16) = 8.64$, $p < .01$, indicated that discrimination was less accurate overall for Japanese than for American listeners. This was due to their poorer performance on pairs at the /w/ end of the continuum (1-4, 2-5) and on cross-category pairs (3-6, 4-7), as indicated by a significant Groups × Comparison Pairs interaction, $F(6,96) = 3.23$, $p = .01$, and simple effects tests of individual pairs. Thus, while both groups showed a single discrimination peak, the Japanese peak was shifted slightly toward the /r/ end of the continuum, and was broader and lower than the American peak. Both of these effects are consistent with cross-language phonemic and phonetic differences, as discussed in the Introduction. The identification test had provided marginal evidence that the Japanese /w-r/ boundary was shifted toward the /r/ end of the continuum, relative to the Americans' boundary, a pattern now corroborated by the small rightward shift of the peak in the Japanese' discrimination function. This shift, although slight, is compatible with the greater cross-language phonetic similarities for /w/ than for /r/. As was argued earlier, the lack of rounding in the Japanese /w/ should lead Japanese listeners to identify more /w/'s in the /w-r/ (as well as the /w-j/) series. Correspondingly, the poor fit of English /r/ to either the Japanese /w/ or the Japanese /r/ categories should converge on perception of fewer /r/'s by the Japanese on the /w-r/ series. English /w-r/ was expected to be assimilated as a category goodness difference within Japanese /w/, English /r/ being heard as a poor Japanese /w/. The lower, broader peak in Japanese discrimination, relative to the American /w-r/ peak and to the Japanese /w-j/ peak(s), is compatible with this hypothesis.

Finally, as previously reported for larger groups (MacKain et al., 1981), results on /r-l/ indicated significant differences between Groups, $F(1,16) = 10.14$, $p < .006$, and between Comparison Pairs, $F(6,96) = 17.74$, $p < .001$, as well as a significant Groups × Comparison Pairs interaction, $F(6,96) = 2.90$, $p < .02$. Japanese subjects discriminated cross-category pairs (3-6, 4-7, 5-8) much more poorly than Americans. This was expected, and is compatible with the hypothesis that Japanese listeners assimilate English /r-l/ as poor exemplars of a single category in their own language. Note also the difference in Japanese performance on /w-r/ versus /r-l/ in Figure 4. Their minimal "peak" in discrimination of the cross-category /r-l/ pairs is clearly lower and broader than their peak in discrimination of /w-r/. This relation is compatible with the hypothesis that /r/ and /l/ are assimilated to a single native category, whereas the /w-r/ contrast constitutes a category goodness difference for Japanese.

Differences in discrimination performance by Experienced and Inexperienced Japanese subgroups were also considered. Overall accuracy across English Experience and Series is shown in Table 4 and Figure 5. Both Japanese subgroups performed relatively well on the /w-j/ series; mean levels were similar to the Americans'. For /w-r/ the Japanese subgroups showed similar performance levels (but note the difference in the position of their performance peaks, Figure 5), although their performance was lower than Americans. Inexperienced Japanese showed lower /r-l/ performance than Experienced Japanese, but again both groups performed less well than Americans.

An English Experience (Americans, Experienced Japanese, Inexperienced Japanese) × Series × Comparison Pairs ANOVA revealed significant effects of English Experience, $F(2,15) = 4.70$, $p < .03$, Series, $F(2,30) = 6.16$, $p < .01$, and Comparison Pairs, $F(6,90) = 25.40$, $p < .01$, as well as significant two-way and three-way interactions [Series × English Experience, $F(4,30) = 3.34$, $p < .03$; Comparison Pairs × English Experience, $F(12, 90) = 1.93$, $p < .05$; Series × Comparison Pair, $F(12, 180) = 6.24$, $p < .001$; Series × Comparison Pair × English Experience, $F(24, 180) = 2.04$, $p < .01$]. Analyses of simple effects for Series within Japanese subgroups showed no significant differences in overall accuracy across series for the Experienced Japanese ($p = .10$), although peaks and troughs were positioned differently across series, as indicated by their significant Series × Comparison Pairs interaction, $F(12,36) = 4.84$, $p < .01$. In contrast, a significant Series effect for the Inexperienced Japanese indicated more accurate discrimination of /w-j/ pairs than of /w-r/ or of /r-l/ pairs, $F(2,8) = 9.31$, $p < .01$. A planned linear contrast on the predicted performance pattern (/w-j/ > /w-r/ > /r-l/) was also significant for the latter subgroup, $F(1,2) = 16.85$, $p < .01$.

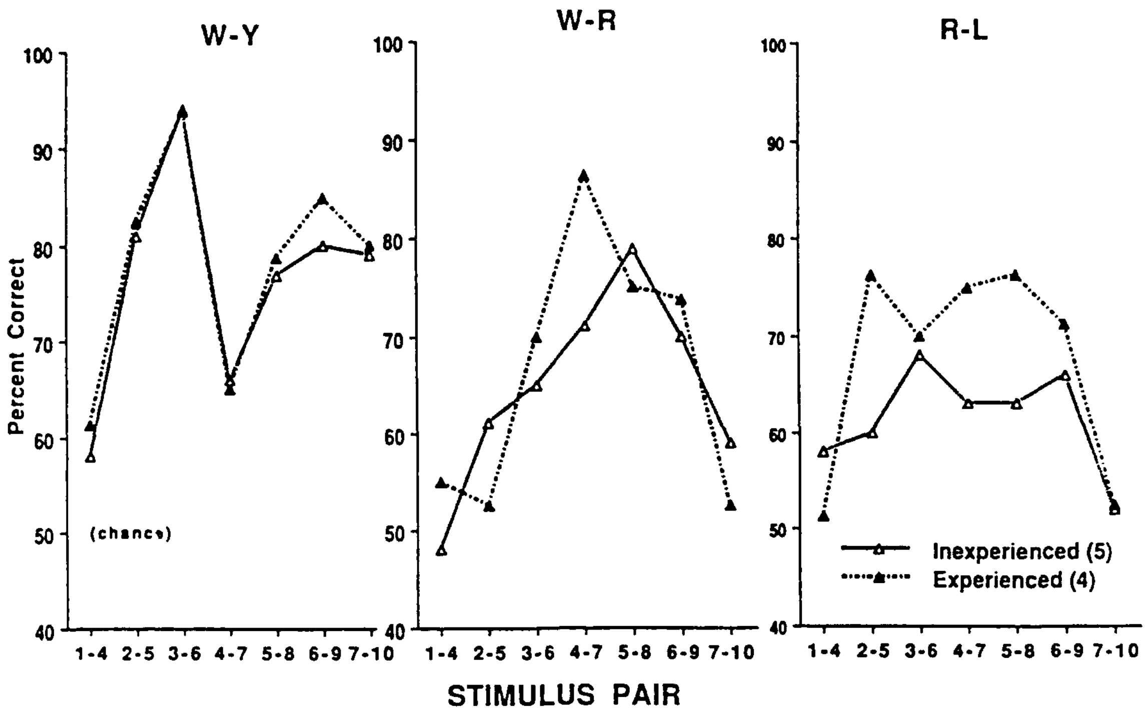

Figure 5. **Average discrimination functions for the Experienced and Inexperienced Japanese subgroups on the three series.**

Experienced and Inexperienced subjects performed almost identically on the /w-j/ series; both groups displayed double peaked functions, which suggest that all the Japanese subjects could differentiate acoustically intermediate stimuli from both /w/ and /j/ phonetic endpoints. An English Experience × Comparison Pairs simple effect ANOVA for /w-j/ revealed no significant effect of English Experience ($p = .66$) and a marginally significant English Experience × Comparison Pairs interaction ($p = .08$). The latter suggests a tendency for the discrimination peaks to be higher, and for the peak between /j/ and the intermediate stimuli to be shifted toward /j/, in both Japanese subgroups relative to the Americans.

There were obvious differences in the pattern of discrimination for Experienced and Inexperienced subgroups on /w-r/ and /r-l/. Separate English Experience × Comparison Pairs analyses revealed significant overall group differences in discrimination of /w-r/, $F(2,15) = 4.16$, $p < .04$) and of /r-l/, $F(2,15) = 5.56$, $p < .02$. Planned linear contrasts indicated that the expected ordering of performance (American > Experienced Japanese > Inexperienced Japanese) was significantly upheld for both series [$F(1,2) = 6.85$, $p < .02$ and $F(1,2) = 10.38$, $p < .01$, respectively]. Performance by the two Japanese subgroups on /w-r/ suggested an effect of experience on the location of the phonetic boundary. This was corroborated by a significant English Experience × Comparison Pairs interaction, $F(6,90) = 2.08$, $p < .04$. While discrimination for Experienced Japanese was most accurate for comparison pair 4-7 (as it was for Americans), the Inexperienced Japanese performed best on pair 5-8. For /r-l/, English experience instead affected the height of the discrimination peak across the category boundary. Consistent with the larger dataset reported in MacKain et al. (1981), Experienced Japanese showed better discrimination than Inexperienced Japanese on cross-category pairs (4-7, 5-8).

2.3 Discussion

Both the identification and the discrimination results are consistent with predictions based on the perceptual assimilation model (Best, 1992; Best et al., 1988). That is, American English /w-r/ appears to be perceived as a category goodness difference within one Japanese phoneme category (/w/), and /r-l/ are perceived as poor examples of a single category. The identification results and the mean discrimination performance levels on /w-j/ are compatible with the hypothesis that the phones are assimilated to two different Japanese categories (but see the qualifications discussed below). Analyses of the two Japanese subgroups

further corroborated predictions. Specifically, Experienced Japanese performed more like Americans than did the Inexperienced Japanese on all series and measures except for discrimination of /w-j/. On that series, there were no cross-language differences (as expected) except for the within-category comparison pairs at the endpoints of the series; this pattern is compatible with language differences in the phonetic properties of /w/.

There was a surprise, however, in the discrimination results for the /w-j/ series. The double peak in discrimination by the Americans and both Japanese subgroups suggested that all listeners may have perceived three rather than two categories along the series, with some category intermediate between /w/ and /j/ perceived in the central portion of the series. This suggests the possibility that the /w-j/ series actually constitutes a combination of a two category distinction for Japanese (/w-j/), along with a category goodness difference within one of those categories. Comparison between the Japanese identification function and their discrimination performance indicates that most of the intermediate category tokens (5-7) were labeled as ambiguous /w/'s. These items were apparently difficult to discriminate *from one another* but easy to discriminate from "good" /w/'s (i.e., items 1-3, consistently labeled as /w/), suggesting a goodness-of-fit distinction within the Japanese /w/ category. Indeed, when the experimenters listened to this synthetic series, several items near the center of the series were perceived as /l/-like. Consistent with this perception, the F1, F2, and F3 onset frequencies and transition patterns in the central stimuli of the /w-j/ series were quite similar to those of the stimuli in the /r-l/ series that were identified by Americans as /l/. The suggestion that the /w-j/ series actually contained three identifiable categories, /w-l-j/, was examined further with a naive group of Americans in Experiment 2.

3. EXPERIMENT 2

3.1 Method

3.1.1 Subjects. As the original subjects were no longer available for testing, nine new native English-speaking American subjects (3 males, 6 females) participated in the study. Seven were graduate students; the other two were faculty members. All reported normal hearing in both ears. Two additional subjects were eliminated from the final sample after testing, when they indicated that they had been diagnosed as learning disabled in childhood. Both had phonemic categorization difficulties, having failed to consistently categorize and discriminate synthetic /ra/-/la/ in a separate but concurrently-run study.

3.1.2 Stimuli and Procedures. The /w-j/ series from Experiment 1 was again employed. The procedure and testing conditions were identical to those of Experiment 1, except that the forced-choice identification test included three response alternatives ("W," "L," "Y") rather than two.

3.2 Results

3.2.1 Identification test. As illustrated in the left side of Figure 6, subjects consistently divided the continuum into three sharply-defined categories. Table 5 lists the means and standard deviations of the boundary location and slope values for both boundaries, computed from PROBIT analyses as in Experiment 1. Three of the 18 fitted ogives deviated significantly from the raw data, according to χ^2 analyses, two on the /l-j/ boundary and a third on the /w-l/ boundary. In all cases, the ogive was the best fit obtainable, and the significant χ^2s were due to extremely steep category boundary slopes.

The location of /w-l/ and /l-j/ boundaries obtained in the three-choice identification task was compared with the /w-j/ boundaries obtained in the two-choice task of Experiment 1. A Groups (Americans-Exp. 2 vs. Americans-Exp. 1 vs. Japanese-Exp.1) × Comparison Pairs ANOVA comparing the /w-l/ boundary with the /w-j/ boundaries yielded a significant main effect of Groups, $F(2,24) = 25.04$, $p < .001$. Sheffe's tests showed that the /w-l/ boundary differed from both the American and Japanese /w-j/ boundaries in Experiment 1 ($p < .01$). In a separate ANOVA comparing the /l-j/ boundary with /w-j/ boundaries from Experiment 1, there was again a significant main effect of Groups, $F(2,24) = 7.86$, $p = .001$. Scheffe's tests indicated that the /l-j/ boundary again differed from the Americans-Exp. 1 /w-j/ boundary ($p < .01$). However, it did not differ from the Japanese /w-j/ boundary ($p = .35$). Thus, while the Experiment 1 discrimination results suggest that the Japanese had actually perceived three categories along the /w-j/ series, as do Americans, the latter result suggests that the Japanese assimilated the intermediate tokens to their /w/ category but as perceptibly poorer exemplars of that category.

Neither the /w-l/ nor the /l-j/ slope values differed from those found for either group in Experiment 1.

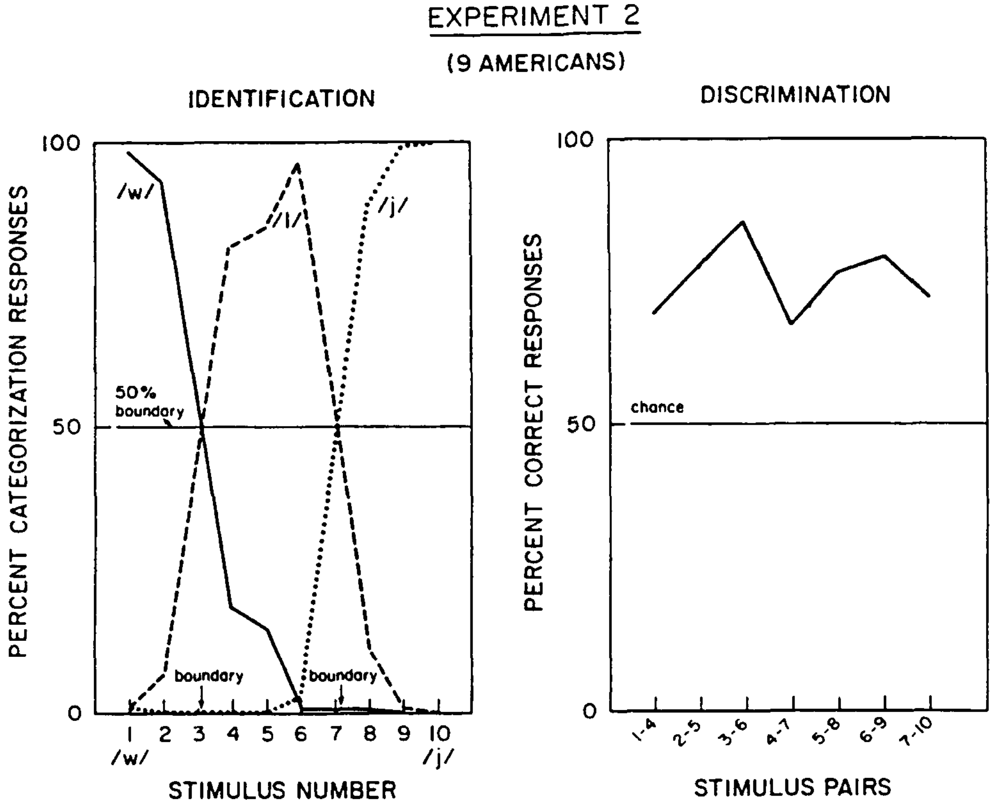

Figure 6. **Identification and discrimination functions for the 3-category tests on the /w-j/ series with Americans in Experiment 2.**

Table 5. *Category boundary locations and slope values for Americans' three-choice identification of the /w-j/ series (Experiment 2).*

	/w-l/ mean	(SD)	/l-j/ mean	(SD)
Boundary Location	3.26	(0.84)	7.20	(0.59)
Boundary Slope	2.53	(1.39)	3.17	(1.17)

3.2.2 Discrimination test. As can be seen in the right side of Figure 6, the discrimination function again showed two peaks of relatively accurate performance, which coincided with the two category boundaries revealed in the 3-choice identification task. For comparison with Experiment 1, a Groups (Japanese-Exp. 1, Americans-Exp. 1, Americans-Exp. 2) × Comparison Pairs ANOVA was conducted. The Groups main effect was non-significant (p = .66), indicating no systematic differences among groups in overall discrimination performance. The significant Comparison Pairs effect, $F(6,144) = 29.68$, $p < .001$, revealed that there were two reliable peaks in discrimination. Finally, the Groups × Comparison Pairs interaction was significant, $F(12,144) = 2.48$, $p < .01$, due primarily to differences among the groups in discrimination of the within-category Pairs (1-4, 3-6, 7-10). However, the locations of discrimination peaks did not differ among the three subject groups.

3.3 Discussion. The results of Experiment 2 confirm that the intermediate category suggested by the double peak in the Experiment 1 discrimination functions was identified by Americans as /l/. As suggested earlier, this categorization is interpretable on the basis of the similarity between the acoustic properties of /l/ and those of the intermediate tokens in the /w-j/ series (see Figure 1). For intermediate tokens, F1 had a steady-state onset, followed by a moderately steep transition, like /l/ but unlike /r/ in the /r-l/ series. They had F2 onsets around 1200-1400 Hz, with a shallow falling transition, again like /l/ in the /r-l/ series. Moreover, their F3 transitions were nearly flat or slightly falling, like that of /l/ in the /r-l/ series, except for a slight dip in frequency just before reaching the vowel steady-

state. In particular, the F3 onset frequency of these stimuli was not close to the frequency of F2, which is needed for good /r/ perception. Given that Japanese does not employ an /l/ phoneme, this intermediate category may have been discriminated from both /w/ and /j/ as a category goodness distinction, most likely within the Japanese /w/ category.

4. General Discussion

The results of Experiment 1 revealed language-specific influences in the perception of English approximant contrasts by adult native speakers of American English and Japanese. Identification and discrimination performance were consistent with cross-language differences in both the phonemic status and the phonetic details of the three contrasts. Both language groups showed sharp category boundaries and high discrimination peaks on the /w-j/ series, which represents a phonemic contrast in both languages. However, there were group differences in the location of the /w-j/ category boundary. The Japanese identified more items as /w/, consistent with cross-language phonetic differences in degree of lip-rounding during production of /w/. On the /w-r/ series, the Japanese showed a more gradual crossover in identification functions and less accurate between-category discrimination than the Americans. In addition, a marginal shift in boundary location and discrimination peak suggested that Japanese categorized more intermediate tokens as /w/ than Americans did. This pattern is also consistent with cross-language differences in the phonetic realization of the /w-r/ contrast. Thus, while in abstract phonological terms /w/ vs. /r/ is a distinctive contrast in Japanese, the phonetic differences across languages led to distinctly different patterns of perception of the synthetic /w-r/ stimuli. As for /r-l/, the Inexperienced Japanese showed much less consistent identification functions and markedly poorer discrimination than the Americans. However, there was no significant shift in boundary location relative to Americans, in keeping with earlier reports (MacKain et al., 1981; Miyawaki et al., 1975). This group difference is compatible with the fact that /r-l/ is a phonemic distinction only in English, and that neither segment is phonetically similar to the Japanese /r/.

This pattern of cross-language differences supports predictions based on the perceptual assimilation model proposed by Best and colleagues (Best, 1992; Best et al., 1988) to explain variations in the difficulty of discriminating nonnative segmental contrasts. Specifically, Japanese listeners were expected to assimilate the English /w-j/ contrast as a two category contrast. The pattern of Japanese listeners' sharp category boundary and high discrimination performance on the /w-j/ series was consistent with this prediction. English /w-r/ was expected to be assimilated to Japanese as a contrast involving a category goodness difference, with /r/ most likely being assimilated as a "poor" exemplar of Japanese /w/. Japanese listeners' more gradually sloping identification function and lower discrimination peak for the /w-r/ series were compatible with this prediction. Finally, English /r-l/ was expected to be assimilated to a single category by Japanese, with both phones representing poor exemplars of either the Japanese /w/ or, less likely, of their tapped /r/. Once again, the more poorly defined category boundary and lower discrimination performance of the Japanese listeners were consistent with this prediction.

The present study extended the model of perceptual assimilation from simple predictions about discriminability of nonnative segmental contrasts to two measures of how nonnative segments are actually categorized by listeners. The location of the category boundary differed between the two groups, consistent with the articulatory-phonetic (and acoustic-phonetic) differences between the American English and the Japanese /w-j/ contrast. Specifically, the Japanese perceived more tokens as /w/ than the Americans, in keeping with observations that Japanese /w/ is more similar to /j/ acoustically and articulatorily than is English /w/. The stimulus items in the /w-j/ series that were identified as /w/ by Japanese but as /j/ by Americans in Experiment 1 were just those items perceived as /l/-like by Americans when they were given a 3-way choice (/w-l-j/) in Experiment 2. Language-specific differences in the phonetic details of the phoneme contrast "shared" by the two languages resulted in a divergence between language groups in the location but not the steepness of the /w-j/ category boundaries across Experiments 1 and 2, which supports the notion that the Japanese listeners assimilated the nonnative segments to the familiar categories of their native phonological system. This language-specific boundary shift extends Lisker & Abramson's (1970) classic findings on cross-language differences in the voice-onset-time boundary for stop consonants to a place-of-articulation distinction for approximants. Moreover, the cross-language differences in identification and discrimination of /w-r/ (and /r-l/)

are quite consistent with differences in the phonemic status and phonetic details of those contrasts with respect to the two languages.

The results of this study are also relevant to Flege's account of cross-language differences in speech perception. According to his Speech Learning Model (1988, 1990) adult learners perceive phones of the L2 on the basis of their "phonetic similarity" to native language (L1) categories. Highly dissimilar phones (referred to as New phones) are initially difficult to categorize perceptually, but with L2 experience, learners form distinct L2 phonetic representations of these categories, which leads to improvement in both their perception and production. Phones which are identical to or highly similar to native phones (Identical phones) are easily perceived even by beginning L2 learners, because they "fit" L1 categories. Phones which are similar to but not identical with L1 categories ('Similar" phones) are the most problematic for L2 learners. They continue to classify Similar phones according to L1 categories even after considerable experience, which leads to continued "accented" production and difficulties perceiving that the L2 phones differ from those of L1. Thus, Flege's model assumes that L2 phones are equated with L1 phonemes in a dichotomous, all-or-none fashion; i.e., they are either fully equated with an L1 phone or fail to be equated to an L2 phone. By comparison, the perceptual assimilation model (Best, 1992) instead assumes that listeners can perceive variations in the goodness of fit of an L2 phone to an L1 phoneme category. The latter assumption is compatible with findings that listeners are sensitive to the category goodness of stimulus variations within a given native category (e.g., Grieser & Kuhl, 1989; Miller & Volaitis, 1989). Also note that Flege's model was developed to address perceived similarities between individual L2 phones and individual L1 phoneme categories, whereas the perceptual assimilation model was developed to address the perception of L2 *contrasts*.

If we extend the Flege model to perception of non-native contrasts between phones, the results of experiment 1 are partially consistent with that model. According to Flege's classification scheme, English /j/ is Identical, /w/ is Similar, and /r/ and /l/ are New phones for Japanese learners of English. Both inexperienced and experienced (re: spoken English) Japanese would thus classify stimuli of the /w/-/j/ contrast according to two Japanese categories, resulting in good identification and discrimination. His model would also predict a shift in the category boundary (relative to Americans), reflecting differences between the Japanese and English /w/. The results of experiment 1 are consistent with both expectations. For the /r-l/ series, inexperienced Japanese would be expected to have considerable difficulty, but experienced Japanese would show improved perception, reflecting the establishment of new phonetic categories. This was indeed the case in Experiment 1. In addition, the fact that the category boundary for experienced Japanese was not different from the Americans' supports the prediction that they had established new L2 categories. However, predictions for the /w-r/ series are somewhat more difficult to generate from Flege's model. The model should predict good identification and discrimination of these stimuli by experienced Japanese, who should have formed a New L2 category for /r/ to contrast with the Similar category of /w/. Their performance levels should therefore equal those of the Americans. However, it is less clear how inexperienced Japanese should perform with /w-r/. Although they would be predicted to identify /w/ well, and /r/ poorly, their discrimination performance is more difficult to predict. Should their performance be poor because they have difficulty with the /r/ that has not yet been established as a New L2 category, or should their performance be moderately good because they perceive /w/ as Similar and recognize that /r/ is different from /w/? In either case, we might expect, nonetheless, that discrimination performance would be lower for inexperienced Japanese than for Americans or for Japanese who are more experienced with spoken English. The shift in discrimination peak for the experienced Japanese toward the location of the American boundary in experiment 1 suggests that those subjects may indeed have established a New /r/ category, which contrasts with the Similar /w/ category. Note, however, that the overall level of discrimination performance did not differ significantly among inexperienced Japanese, experienced Japanese and Americans, as would be predicted from Flege's model.

Flege's model might also appear to address the existence of the intermediate category in the /w-j/ series, even for Japanese listeners, i.e., they may have begun to form a new /l/ category as a result of English experience. However, two observations are at odds with this possibility. First, there was no difference on that contrast between the Inexperienced Japanese, who had had very little experience with spoken American English at the time of testing, and the Experienced Japanese.

Both groups provided equally strong evidence of perceiving the intermediate category in the /w-j/ series; the intermediate category in the double-peaked discrimination functions was no less clear for the Inexperienced Japanese than for the Experienced Japanese, or in fact for the Americans. Second, if even the Inexperienced Japanese were truly developing a new phonetic category on the basis of their limited English exposure, then we would expect this /l/ category to emerge in their responses to the /r-l/ series as well. Such was not the case.

Flege's notion that L2 experience may lead to the formation of new phonetic categories is not incompatible with Best's perceptual assimilation model. The assumption that experience with spoken L2 may lead to a reorganization of perceptual assimilation of nonnative phones, in fact, motivated the comparison between the Japanese subgroups differing in English conversation training and experience. The assimilation model assumes that listeners are sensitive to degrees of similarity and dissimilarity between the nonnative and native phones. This is most obvious when there are category goodness differences in assimilation, or when the nonnative phones are non-assimilable. Indeed, adult L2 learners should be expected to form new phonetic categories most readily for L2 phones perceived as discrepant exemplars of a native category, i.e., for the non-prototypical member of a contrast that is assimilated as a category goodness difference from a native phoneme. If no discrepancies are perceived between the L2 and L1 phone—that is, for the L2 phone that is perceived as a good exemplar of the native phoneme—it should be quite difficult for the L2 learner to form a new category. Conversely, if the L2 phone is so *dis*similar from L1 phonemes that it cannot readily be related to any L1 category, we may expect the L2 learner to have some difficulty forming a new phonetic category, because a clear *contrast* between a specific familiar phoneme and an unfamiliar phone may be particularly informative to the learner.

The one unexpected finding—that listeners from both language groups apparently discriminated a third, intermediate phonetic category between the two endpoint categories of the /w-j/ series—is consistent with the above suggestion. Experiment 2 with a new group of American listeners verified that this third category was highly identifiable as /l/ (although it remains to be determined whether Japanese listeners at either level of English experience would reliably label those items as "L"). Although the Japanese language does not employ an /l/ phoneme, even the Inexperienced Japanese clearly distinguished a third phonetic category from the /w/ and /j/, according to the two marked peaks in their /w-j/ discrimination function, which was virtually identical to the discrimination functions of the two groups of Americans. This observation, together with the Inexperienced Japanese listeners' better discrimination performance on /w-r/ than on /r-l/, suggests the possibility that adults' recognition of the phonetic properties of a nonnative segment might be aided by direct comparison between exemplars of that segment presented in context with exemplars of the most similar (in articulatory-phonetic or acoustic-phonetic terms) native phoneme. That is, perceptual learning about the novel L2 segment may benefit from contextual comparisons which exemplify differences between the native phoneme and the nonnative phone that is perceived as a poorer exemplar of that familiar category. In the present context, Japanese listeners' recognition of a third category in the /w-j/ series, which was identified as /l/ by the Americans in Experiment 2, apparently benefited from its contrast to the flanking categories of Japanese /w/ and /j/, i.e., the intermediate, nonnative category constituted a noticeably poor fit to one or both of the familiar Japanese categories. While this observation is consistent with Flege's (1988; 1990) claim about the importance of similarity versus "newness" of nonnative phones to the degree of perceptual adjustments to L2 learning, it is also compatible with the perceptual assimilation hypothesis that category goodness differences are relatively discriminable as a difference between the native category "ideal" and less-good exemplars. Further research is obviously needed to determine whether presenting a nonnative phone in juxtaposition to the most similar native phoneme contrast may actually improve perception of the new category.

In either event, the data presented here are generally consistent with the suggestion that language-specific attunement of phonetic perception may remain somewhat malleable even in adulthood (see also Flege, 1988; MacKain et al., 1981; Pisoni et al., 1982; Strange & Dittmann, 1984; Tees & Werker, 1984; Werker & Tees, 1984). The subgroup of Japanese listeners who had had more intensive conversation experience with American English speakers showed greater similarities to the Americans than did the Inexperienced Japanese in their performance on all three stimulus series. Thus, English conversation experience may have shifted those Japanese listeners' catego-

rization and discrimination toward the phonemic and phonetic properties of the approximant contrasts employed in American English. Note, however, that the performance of the Experienced Japanese was not identical to the Americans', instead falling intermediate between the latter group and the Inexperienced Japanese (see also Yamada & Tokhura, 1991).

Further research is needed to determine which factors may influence adults' perceptual adjustments to the phonemic and phonetic properties of L2 segmental contrasts, and to what extent there may be limitations on such L2 influences in adulthood. It is important to recognize that we had no control over, or access to, the factors that led to the group differences in English conversation experience. For example, in our Japanese subgroups, level of English conversation experience may have been affected by individual differences in phonetic ability (recall the categorical /r-l/ performance of the Inexperienced Japanese subject M. K.: MacKain et al., 1981), by differences in the necessity of speaking English, by differences in motivation to use English "like a native," and/or by differences in the nature of exposure to English (e.g., traditional classroom vs. immersion program), in addition to duration and intensity of exposure to spoken English. Another factor that appears to have strong impact on an adult's ability to perceive a given nonnative contrast is whether the individual had any substantive exposure during early childhood to languages using that contrast (e.g., Flege, 1988; Tees & Werker, 1984).

Although we cannot verify that the Japanese subgroup difference we found was due to differences in L2 experience in adulthood, rather than to earlier-occurring factors, several observations suggest the likelihood that the relevant experience with spoken L2 was limited to adulthood. Three of the Experienced Japanese had come to live in the U. S. as adults, the fourth at 19 years, all past the presumed "critical period" for language-learning which ends at puberty. All had begun intensive English conversation training either after their arrival in the U. S. or less than a year before they left Japan. Moreover, while most Japanese are formally taught English in school beginning at age 12 years or earlier, the instructors are typically native Japanese rather than English speakers, and the emphasis is on reading/writing and not on speaking/hearing (Mochizuki, 1981; Yamada & Tokhura, 1991). Nonetheless, further research is needed to clarify the contribution of various factors to subgroup differences in perception of L2 contrasts, including studies of longitudinal changes within a given group of listeners.

REFERENCES

Best, C. T. (1992). The emergence of language-specific phonemic influences in infant speech perception. In H. C. Nusbaum & J. Goodman (Eds.), *The transition from speech sounds to spoken words: The development of speech perception*. Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29*, 191-211.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*, 86-125.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 2.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62*, 961-970.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (Ed.), *Human communication and its disorders: A review*. Norwood, NJ: Ablex.

Flege, J. E. (1990). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Huebner (Eds.), *Crosscurrents in second language acquisition and linguistic theories*. Philadelphia: John Benjamins.

Flege, J. E., & Bohn, O.-S. (1989). The perception of English vowels by native Spanish speakers. *Journal of the Acoustical Society of America, 85 (Suppl.)*, S85(A).

Gillette, S. (1980). Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language, 6*, 59-72.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14*, 339–342.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9*, 317-323.

Grieser, D. L., & Kuhl, P. K. (1989). The categorization of speech by infants: Support for speech sound prototypes. *Developmental Psychology, 25*, 577-588.

Kasuya, H., Takeuchi, S., Sato, S. & Kido, K. (1982). Articulatory parameters for the perception of bilabials. *Phonetica, 39*, 61-72.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia.

Lisker, L. (1957). Minimal cues for separating /w,j,r,l/ in intervocalic position. *Word, 13*, 256-267.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*, 505-512.

Miyawaki, K. (1973). *A study of lingual articulation by use of dynamic palatography*. Unpublished masters thesis, Department of Linguistics, University of Tokyo.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic

experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9*, 283-303.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w,j,r,l/ in English. *Word, 13*, 24-43.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 297-314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Pollack, I., & Pisoni, D. B. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science, 24*, 299-300.

Price, P. J. (1981). *A cross-linguistic study of flaps in Japanese and in American English*. Unpublished doctoral dissertation, University of Pennsylvania.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts. *Journal of the Acoustical Society of America, 87 (Suppl.)*, S72(A)

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243-261.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131-145.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. *Canadian Journal of Psychology, 38*, 579-590.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development, 47*, 466-472.

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*, 35-44.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866-1878.

Yamada, R. A., & Tokhura, Y. (1991). Age effects on acquisition of non-native phonemes: Perception of English /r/ and /l/ for native speakers of Japanese. *Proceedings of the 12th International Congress of Phonetic Sciences, 4*, 450-453.

FOOTNOTES

*Journal of Phonetics, 20, 305-330. (1992).

†Also Wesleyan University.

††University of South Florida, Tampa.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 109-118

Plausibility, Parsimony, and Theories of Speech*

Alvin M. Liberman

According to a somewhat unconventional view, speech is managed by a specialization for language—a phonetic module—at the level of action and perception. There, the processes and primitives are specifically phonetic, not, as is more commonly assumed, generally motor and auditory. The less conventional view is nevertheless the more plausible because it (1) better illuminates the biological nature of the difference between spoken and written forms of language, and (2) provides the better account of how speech meets the specific requirement of phonological communication that the elements be commutable, as well as the general requirement of all communication systems that there be parity between sender and receiver. Also relevant to the argument of plausibility is the fact that, while the phonetic module is unique to language, it is not without biological precedent, since it has important properties in common with such older (and better understood) specializations as stereopsis and sound localization.

It is, for me, a happy privilege to be part of an occasion that honors Paul Bertelson, dear friend and valued colleague. As my contribution to the occasion, I offer a few reflections on a question I have often discussed with Paul: Is there a specialization for language at the precognitive level? Is there, in other words, a specifically linguistic mode of action and perception? Put in one form or another, this question goes to the heart of claims about the modular nature of linguistic processes. It arises wherever in language one happens to look, but it assumes what I take to be its most pointed manifestation at the level of phonetic structure. There lie two or three dozen consonants and vowels, familiar objects of a seemingly simple sort. Yet they are the elements of which all languages are made. Moreover, their proper use is a distinguishing mark of the human species and a principal component of its linguistic faculty. Accordingly, the question I raise about their management is a question about the biology of language.

Together with some of my colleagues, including especially Ignatius Mattingly, I believe the answer to the question is yes—the biology of language does, indeed, incorporate a precognitive specialization for the production and perception of consonants and vowels, a specialization we have chosen to call a phonetic module. We take this module to be an integral part of the larger specialization for language, adopting what Fodor (1983) would characterize as a *vertical* view in which the relevant structures and processes are seen as specific to the linguistic function they serve. The opposite view, which is more widely held, is that speech is to be accounted for by the most general principles of motor activity and auditory perception; accordingly, this view is appropriately referred to as *horizontal*.

My aim in this paper is to promote the less conventional vertical view, not by reference to the results of particular and putatively critical experiments, but rather by taking account, in very general form, of a few commonly neglected considerations that are relevant to its plausibility and parsimony. A fuller description of the vertical view, together with an account of the nature of its empirical support, is to be found elsewhere (Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Mattingly & Liberman, 1988). As will be seen there, this view comprehends both the production and perception of speech; indeed, it assumes an organic relation between the two. It happens, however, that the considerations I mean to offer in this paper are concerned primarily with perception, so I will bias the emphasis in that

This work was supported by NIH grant HD-01994 to Haskins Laboratories.

direction, as I do in the following brief account of the difference between the vertical view and its horizontal opposite.

The horizontal view varies in its particulars from one theorist to another, but the basic assumptions are much the same. Thus, the several proponents are in agreement that perception of speech is no different from perception of other sounds (Ades, 1977; Bregman, 1991; Cole & Scott, 1974; Crowder & Morton, 1969; Diehl & Kluender, 1989a; Fujisaki & Kawashima, 1970; Howell & Rosen, 1984; Kuhl, 1981; Lane, 1965; Lindblom, 1991; Miller, 1977; Oden & Massaro, 1978; Stevens, 1981). All such perception is supposed to depend on the same general processes of hearing, processes that occupy a common domain and evoke in a common sensory register a common set of auditory primitives, including, for example, pitch, loudness, and timbre. Of course, the perceptual representations of a stop consonant and, say, a squeaking door must be different, but the difference is supposed to be only in the relative values that are assigned to the primitives they have in common; there are no specifically phonetic primitives. Thus, the primary perceptual representations of speech are taken to be generally auditory, not specifically phonetic. That being so, proponents of the horizontal view are required to explain how, being independent of language, the auditory representations gain access to a system in which they are specifically marked for linguistic significance and used for a specifically linguistic purpose.

Some proponents explicitly meet this requirement by supposing that, given the auditory percepts, the listener elevates them to linguistic status by attaching phonetic labels, fitting them to phonetic prototypes, or associating them with such cognitive units as *distinctive features* (Ades, 1977; Crowder & Morton, 1969; Fujisaki & Kawashima, 1970; Pisoni, 1973; Rosen & Howell, 1987; Stevens, 1975, 1989). Since these labels, prototypes, and features are neither acts nor percepts, they deserve to be called ideas. But whatever they are called, they are the end products of a cognitive translation that converts auditory percepts into a form appropriate to language. Getting from speech signal to the primary level of language is, therefore, a two-stage process: evocation of an auditory percept in the first stage, followed by conversion to a phonetic representation in the second. In this important respect, the horizontal view implausibly makes perceiving speech no different in principle from perceiving Morse code or, for that matter, the letters of the alphabet; in all cases, the perceiver must attribute linguistic significance to percepts that are not inherently linguistic (see Liberman, in press, for further discussion).

There are at least two other assumptions of the horizontal view, but these are commonly left unsaid, though they are, to the vertical theorist, of great importance. One, which seems to be tacitly accepted, not as an assumption but as background fact, is that phonetic elements are sounds. The other, which is commonly unspoken because it must appear on this view to be irrelevant, is that the gestures and motor control processes of speech production are, like the processes of speech perception, independent of language. Presumably, language simply appropriated movements and motor mechanisms that are part of a general faculty for action, just as it appropriated for its own special purposes the general mechanisms of audition. It is, therefore, necessary for the speaker, just as it is for the listener, to make a cognitive translation between two very different kinds of representations, one linguistic, the other not. According to the horizontal view, then, it should not matter in this regard whether one produces language by speaking it, by operating a Morse-code key, or by wielding a pen. Putting this observation about production together with the earlier one about perception, we see that the horizontal view must fail in both domains to provide a plausible basis for distinguishing the biologically primary processes of speech from their obviously secondary extensions.

The vertical view is different at all points. Seen vertically, apprehending phonetic structures is managed by a distinct, language-specific system that has its own phonetic domain, its own phonetic mode of signal processing, and its own phonetic primitives. Perception of phonetic structure is therefore precognitive, which is to say immediate; there is no translation from a nonphonetic (auditory) representation because there is no such representation. It is, of course, in precisely this respect that perception of speech differs, plausibly, from perception of Morse code or of scripts.

There are two other assumptions of the vertical view that contrast starkly with its conventional counterpart. One is that the elements of phonetic structure are gestures, not the sounds those gestures produce. These acts are, then, the ultimate constituents of language, the primitives that must be exchanged between speaker and listener if communication by language is to occur. The second assumption is that these gestures, as well as

the processes that control them, are specifically phonetic, having evolved for phonological communication and for nothing else. Unlike a Morse code operator or a writer, a speaker is directly using motor representations that are inherently linguistic. There is no need to connect a nonlinguistic act (pressing a key or writing an alphabetic character) to some linguistic unit of a cognitive sort. Nor is such a unit required, more generally, to serve as a common referent through which a nonlinguistic act and a correspondingly nonlinguistic (auditory) percept can be connected to each other. On the vertical view, the specifically phonetic gestures that are managed by the module in production are recovered by the module as the specifically phonetic primitives of perception, thereby completing the communicative link without cognitive intervention, while also making speech an integral part of language, not, as on the horizontal view, an artifactual adjunct.

Are there acoustic substitutes for speech?

According to the horizontal view, speech percepts are supposed to be auditory in the same way that the percepts evoked by the letters of the alphabet are known to be visual. In the visual case, the only limit to the number and variety of optical shapes that can be made to serve as alphabetic characters is in the constraints imposed by the visual system, and they are few. Given the conventional view of speech, one would suppose that a similar situation would exist there. Of course, the auditory channel is neither so wide nor so deep as the visual, but, still, the number of sounds that can be identified is very great, so one should expect that it would be possible, even easy, to find alternative acoustic vehicles.

The foregoing implication of the horizontal view is exactly what my colleagues and I tacitly accepted when, in 1945, we were enlisted in an attempt to build a device that would convert print into intelligible sound and so serve as a reading machine for the blind. We should, of course, have wanted a machine that would make the print speak English, but there were at the time no such things as optical character readers, and, even if there had been, we should not have known how to synthesize speech from their outputs. However, we considered this to be of no great consequence, for we could quite easily make the print control the parameters of various nonspeech sounds, and so produce an acoustic cipher differing only in detail from the speech to which the blind users were accustomed. Given our tacit assumptions about the nature of speech, we supposed that they would learn to connect these sounds to phonetic units, much as they had earlier done with the sounds of speech.

A detailed account of our unsuccessful attempts to substitute nonspeech sounds for the sounds of speech would not be enlightening here, for it would only make the point that, try as we might, we did not come anywhere near to succeeding. Of course, we could not then, and cannot now, expect to test all possible sounds, nor could we readily arrange for people to have with nonspeech the amount of experience they must have had with speech. Still, we were then, as we are now, convinced that nonspeech sounds simply won't do, not just because they failed the tests we put them to, but because they failed in ways that made it plain why we should never have expected them to succeed. The difficulty was not primarily that the sounds were indiscriminable or unidentifiable, but rather that every arrangement we tried was defeated in one way or another by the variable of rate. Thus, we found that, as the rate of scan approached the lower bound of what would be even marginally acceptable in speech or in reading, performance (as measured by ability to learn a selected set of words) decreased appreciably. Worse yet, listeners lost the ability to identify the individual letter sounds and to apprehend their order, responding instead to some overall auditory pattern characteristic of the word. Thus, to the extent that the words could be learned at all, they had to be treated logophonically, as it were, with attention directed to the way the sound differed holistically from the sound for any other word. The tremendous advantage of the combinatorial principle that phonology exploits was therefore lost, and, given that a purely logographic system cannot really work very well even in reading (De Francis, 1989; Mattingly, 1991), one can imagine how vastly more unsuited it would be as a basis for speech perception.

The final blow was dealt by our observation that when we ourselves undertook to master one of these nonspeech systems, we found little transfer of training across rates. Letters and words learned at one rate could not be recognized at other rates that were still within the range of what was reasonable if the machine was to have any utility. Words tended not only to become hard-to-analyze wholes, but the phenomenal nature of the whole changed quite drastically from one rate to another. A user would have been required, therefore, to learn a different set of associations for every significantly different rate.

In hindsight, it is apparent that if we had ever bothered to think about the requirements of phonological communication, and then measured these against the known properties of the ear, we should have realized, without any research at all, that an acoustic-auditory strategy of the kind suggested by the horizontal view was bound to fail. The point is that phonological communication requires commutable, hence discrete and invariant, representations. But if such invariance is to exist in the auditory domain, as it must on the view that we had unthinkingly adopted, then rates of transmission that are normal in speech would seriously strain and sometimes overreach the temporal resolving power of the ear and also its ability to perceive the order of the segments (Liberman, Cooper, & Studdert-Kennedy, 1968). (Speech production would be equally problematic, since invariant and discrete auditory percepts would require correspondingly invariant and discrete gestures, with the result that people could not really speak, they could only spell).

But we had to learn the hard way, as it were, that nonspeech sounds—that is, sounds that do not approximate the results of linguistically significant gestures—cannot be efficient vehicles for language. It was, indeed, this painfully-arrived-at conclusion that initially motivated Frank Cooper and me to begin our speech research. Our aim, very simply, was to find out why the sounds of speech, but no others, can meet the commutability and rate requirements of phonological communication. The answer our research brought us to seems to me now so plausible, not to say obvious, that I wonder we did not arrive at it earlier, simply by thinking about the matter. For what it comes to is that evolution did not ever confront the problems of commutability and rate, simply because it avoided the acoustic-auditory strategy (of the horizontal view) that would have given rise to them. What evolved was a brilliantly successful strategy that defined the invariant elements of phonetic structure not as sounds, but as gestures. The critically important advantage of this strategy was that, given gestures that can somehow be characterized as remote structures of motor control, and given a mode of action specifically adapted to matching these to the needs of phonology, it was possible by overlapping and merging (that is, coarticulation) of the peripheral movements to achieve the high rates of production that characterize speech communication.

As for perception, which was initially our single-minded concern, the advantage is that coarticulation effects parallel transmission of information about successive phonetic segments, and so relaxes the constraints on rate of perception that underlay the failure of our nonspeech reading machines. But this gain has an obvious cost, for coarticulation creates a complex relation between signal and message, a specifically phonetic code that is opaque except as the scientist or perceiving device can take account of the phonetically specific processes that produced it. Once research on speech had convinced us that this was so, we felt challenged to explain, if only in the most general terms, how listeners manage. We rejected the possibility that they *break* the code by some deliberate, cognitive process, preferring, instead, to suppose that they rely on a biologically coherent module specifically adapted to providing the articulatory key. But whatever the plausibility of this proposed solution, it was never plausible to suppose that perception of linguistic structure is so much controlled by general auditory processes that it can be achieved as well with sounds other than speech. That we nevertheless thought it was is testimony to the unquestioning faith we had in what was then, and is now, the received view.

Whence comes the fit of perceptual form to phonological function?

Given that the function of phonology is to use the combinatorial principle to generate a large number of words, the units must, as already noted, be discrete and invariant, which is to say categorical, as they are seen from a linguistic point of view. It is adaptive therefore that the units be correspondingly categorical in immediate perception. Listeners would only be disconcerted by the sense, if it should be their sense, that a particular phonetic token, X, lay half way between X and Y, or that it really sounded like Z, except as it was reinterpreted so as to take account of the fact that it was followed by A. Fortunately, listeners do not have either sense: the much-investigated peaks of discriminability at the acoustic boundaries of the phonetic unit reflect category-producing discontinuities in perception, and it is characteristic of phonetic perception that these categories remain stable across all context-conditioned variation in the stimulus.

What, then, is the source of these stable perceptual categories? On the horizontal view, it must, of course, be in the properties of the auditory system. Accordingly, theorists of this persuasion take comfort in the experiments that

find categories in the responses of nonhuman animals to speech and in the responses of human listeners to acoustic nonspeech analogues (Diehl & Walsh, 1989; Kluender, 1991; Kluender, Diehl, & Killeen, 1987; Kluender, Diehl, & Wright, 1988; Kuhl & Miller, 1975; Massaro, 1987; Parker, 1988; Parker, Diehl, & Kluender, 1986; Pastore, 1987; Pisoni, 1973; Pisoni, Carrell, & Gans, 1983). The opposite result is also found, much to the satisfaction of the vertical theorists, who must believe that this kind of categorical perception is specifically phonetic (Best, Morrongiello, & Robson, 1981; Best, Studdert-Kennedy, Manuel, Rubin-Spitz, 1989; Mann & Liberman, 1983; Liberman, Isenberg, & Rakerd, 1981; Mattingly, Liberman, Syrdal, & Halwes, 1971; Sinnott, 1976; Waters & Wilson, 1976). However, I do not mean here to offer a critical evaluation of the experimental evidence pro and con the one assumption or the other, but, rather, in keeping with the spirit of this paper, to argue that the horizontal (auditory) interpretation is simply implausible on its face.

It is relevant, first, to take into account how very great is the variation in stimulus for any given perceptual category (Repp & Liberman, 1987). For all phones, there is variation as a function of phonetic context, position in the syllable, and vocal-tract size. In some cases, there are changes depending on articulatory rate and stress. And, of course, there are the differences that exist across languages. Indeed, so gross is this stimulus variation, and so numerous its sources, that it is impossible to estimate how very many alternative category boundaries the auditory system would need if the percepts were to be held constant, and implausible to suppose that these boundaries could exist in such numbers. Surely, they could not have been selected in the evolution of the auditory system just against the possibility that phonology would one day come along and find them useful. Yet, as properties of the auditory system, they serve no other imaginable purpose. Indeed, from an auditory standpoint, they would be dysfunctional, since they would necessarily distort the perception of nonspeech sounds.

Even if one assumes, against all reason, that this numerous variety of boundaries does exist in the auditory system, is it plausible to suppose that coarticulatory maneuvers vary as they do with phonetic context and with rate just in order to produce sounds that match the way categories of the auditory system happen, independently of coarticulation, to adjust to variation in the acoustic stimulus?

Moving, now, from implausibility to impossibility, I remark the fact that, as is well known, the articulation of every phonetic unit has multiple acoustic consequences, and that listeners are more or less sensitive to all of them. So, if speakers had somehow managed to produce a second-formant transition to fit some auditory category, what then would they do about the third-formant transition and the burst? The answer has got to be nothing, since it is not possible to control these acoustic consequences independently.

It is also true of these multiple sources of information that, no matter how numerous and acoustically various they may be, they nevertheless evoke a unitary, categorical percept. This equivalence of the acoustically very different components of the speech signal is reflected in, and measured by, the trading relations, so-called, that speech researchers report (Diehl & Kluender, 1989; Fitch, Halwes, Erickson, & Liberman, 1980; Repp, 1982). But one hardly needs experiments like those to make the point. For, surely, there is no doubt that there are multiple and acoustically very different sources of acoustic information for every phone, and it is common experience that the result is a unitary perceptual category, not a collage in which the several fragments represent the disparate auditory consequences of the different acoustic cues. Is it even conceivable that speakers produce these heterogeneous combinations of sounds by design, and that they do so because they once discovered that the auditory system just happens to cause them to evoke the same percept. It would, again, be dysfunctional if the auditory system did that, for it would effectively prevent the discrimination (or identification) of most ordinary acoustic events; indeed, it would tend to make all of them sound like speech.

Nor can one reasonably suppose that such categories as the auditory system apparently does have might somehow have served as starting points for the development of phonetic perception (Kuhl, 1981). Which contexts, rates, vocal-tract sizes, and languages might have been taken as the linguistic canon? And even if these auditory categories are appropriate in some phonetic circumstances, would they not be inappropriate, hence dysfunctional, in all others? Indeed, auditory categories, to the extent that they exist, should make us the more convinced of the validity of the vertical view, since they require of the phonetic system

that it be so independent as to ignore their potentially interfering representations.

Is it not far more plausible to suppose about all these cases that the variable and multiple sources of information in the speech signal are simply the inevitable consequences of acts that are specifically adapted to a phonological function, and that perception is managed by a corresponding adaptation to those same acts and that same function?

What is the place of speech in the biological scheme of things?

If, as the horizontal view would have it, there is no specialization for language at the level of action and perception, then, as I have already implied, language must begin one step up, where, by a purely cognitive process, a select set of nonlinguistic representations is given a phonetic cast and so made appropriate for whatever specialized language processing the theorist wishes to assume. The same conclusion follows if the theorist should, by a seemingly logical extension, embrace the more broadly horizontal assumption that there is no specifically linguistic process at any level, that just as speech is merely one among many expressions of the general faculties of action and perception, so does syntax fall out of a general faculty of cognition. On either version, however, it will be hard to provide a parsimonious answer to a fundamental question about the biology of speech: how are the acts and percepts of speech marked in evolution for linguistic significance, and so set apart from all others?

Perhaps the most explicit attempt to answer this question from a horizontal point of view has been made by Lindblom (1991) who says that "languages make their selection of phonetic gesture inventories under the strong influence of motor and perceptual constraints that are language independent and in no way special to speech (the functional adaptation of phonetic gestures)." Then, referring to the unconventional assumption that there are specializations at the level of perception and action, he says, "If so, why do inventories of vowels and consonants show evidence of being optimized with respect to motor and perceptual limitations that must be regarded as biologically general and not at all special to speaking and listening?"

As a criticism of the vertical view, which is how it was intended, Lindblom's argument can be dismissed as irrelevant to the question that this view is designed to answer. That question is not whether language somehow evolved out of what was already there, for it could hardly have done otherwise, but, rather, what it was that evolved. Lindblom's answer is that there was, at the precognitive level, no evolution of anything, only a selection from among the possibilities offered by general faculties that were, and presumably still are, independent of language. Of course, that must have been exactly what happened in the development of, say, a cursive writing system, for surely the selection of its characters must have been strongly influenced by "motor and perceptual constraints that are language independent." But such an observation, true though it is, enlightens us not at all about the evolution of language, for what developed in the case of cursive writing were artifacts, not the biologically primary units of the language that those artifacts are taken to represent. Obviously, the artifacts can have been marked for linguistic significance only by agreement, not by the processes of biological evolution. It is up to each user, then, to honor the agreement by mastering, at a cognitive level, the wholly arbitrary connection between the selected characters and the primary units of the language. On Lindblom's account, the same must be said of speech and the speaker-listener. For if speech production and perception are not distinctly linguistic, the primary units of language must, as earlier noted, be in the nature of ideas—i.e., the labels, prototypes, distinctive features, etc.—to which the nonlinguistic representations of speech become connected. Such ideas might have been a result of the inventiveness that large brains and cognitive power make possible, in which case, the biology of speech would be the biology of large brains and cognitive power. Or, alternatively, they might have become part of the genetic inheritance of human beings, in which case the biology of speech would be the biology of innate ideas. In neither case would there be a place for speech in the biology of language.

According to the vertical view, the biology of speech embraces specifically phonetic structures and processes that are adapted to specific linguistic functions. What evolved, on this view, was a special mode of communication (the phonological mode), that serves a distinctly linguistic function (the generation of a large vocabulary by use of the combinatorial principle), and imposes phonology-specific requirements (among which are the rapid production and perception of commutable elements). The primitives of this mode are correspondingly special, being specifically linguistic and so appropriate for their role in the larger specialization for language, including, for example,

the syntactic component. On that basis, it seems plausible to suppose that the elements and processes of the phonological mode were selected according to their ability to meet its special requirements. On the side of action, I should think that an important factor was not ease of production as such, but rather the extent to which the gestures lent themselves to the coarticulatory maneuvers that effectively circumvent the constraints on rate that would have been imposed had discrete gestures been produced seriatim. On the perceptual side, a decisive factor must have been the immense advantage conferred by a complex kind of parallel transmission that extends the limit on rate set by the temporal resolving power of the ear. It would appear then that, so far from being driven to exploit the strengths of the general motor and auditory systems, as Lindblom's comments imply, the evolution of speech must have been guided, rather, by the need to find ways around what must be seen, from a phonological point of view, as their weaknesses. It must also have been guided, even more generally, by the need to meet the requirement of parity by establishing an identity between the communicative acts of the speaker and the communicative percepts of the listener. This it did by incorporating in the precognitive biology of speech the special mechanisms that allow articulatory gestures—the constituents of language that must be common to speaker and listener—to survive the rigors of the communicative exchange.

It is also relevant to the plausibility of a theory of speech to expose, among its biological implications, the relation of speech to other forms of natural communication. On any theory, the gulf between speech and other systems must, of course, be seen to be very wide, though one would surely be inclined to look with favor on a theory that nevertheless managed some kind of bridge. It therefore counts against the horizontal view that it fails to do that. For if there is no precognitive specialization for speech, then, as has been noted several times already, speech must be matched to phonetic ideas. The horizontal theorists apparently find that consequence acceptable as it applies to human beings and their language. But would they not hesitate to extend it to the nonhuman case? Presumably, they would, given the abundant evidence that nonhuman communication is underlain by specializations for producing and perceiving specifically communicative signals of one sort or another. Are we to suppose, then, that unlike the nonhuman animals, which communicate as they do because of the nature of their precognitive specializations, we humans speak because, having risen above that mean level, we take advantage of innate ideas and intelligence? The vertical view, on the other hand, permits us to see that we and the other creatures are all precognitively specialized for communication; the important difference is that our specialization comprises a phonology and a syntax, while theirs does not.

There remains the biologically relevant question: What more general phenomena are exemplified by the processes of speech? Here, the horizontal view might appear to have the advantage, since it takes speech production and perception to be not different from other forms of action and perception. Accordingly, speech processes are as general as those that manage all of auditory perception and all of motor activity. The vertical view, on the other hand, abjures this kind of generality, holding that speech processes are specific to the linguistic function they serve. Indeed, it is precisely on this score that the unconventional view has been criticized as unparsimonious. As I have already tried to show, however, it is just because of the assumption about special processes that the unconventional view is the more parsimonious, since assuming another precognitive specialization is presumably less in need of Occam's razor than assuming a set of innate phonetic ideas.

At all events, assuming a specialization for speech is no more unparsimonious than making the corresponding assumption for other systems that are biologically adapted to stimulus events and properties that are of great ecological significance to the species. Consider, for example, echolocation in the bat, sound localization in the barn owl, song in the bird, or, indeed, stereopsis in the human. Like the speech specialization as characterized by the vertical view, each of these is to be understood only by reference to the special mechanisms by which it serves its special function. While each system is therefore different from every other, they have in common the properties that Fodor has identified as characteristic of the modules that he takes as the functional elements of the precognitive mind. Moreover, the specializations named above have in common with each other and with speech that they all belong to a class of modules called 'closed' by Mattingly and me, and claimed by us to share the following properties (Liberman & Mattingly, 1989; Mattingly & Liberman, 1988).

(1) The representations are heteromorphic. That is, the dimensions of the percept are in-

commensurate with the dimensions of the stimulus. Thus, in stereoscopic vision, the viewer perceives heteromorphic depth, not homomorphic disparity (doubling of images). In speech, the listener perceives, heteromorphically, a string of discrete consonants and vowels, not the continuously varying timbres (chirps, whistles, bleats, etc.) that constitute the homomorphic representations of the continuously changing formant tracks.

(2) The modules preempt the stimulus information that is of interest to them, using it to form the heteromorphic percept, while leaving none for the homomorphic counterpart (Bentin & Mann, 1990; Liberman & Mattingly, 1989; Whalen & Liberman, 1987). Thus, over a range of binocular disparities, the viewer perceives depth; disparity is not also seen. In a similar way, listeners perceive phonetic structures, not phonetic structures and also the homomorphic chirps and whistles that the components of the acoustic signal would otherwise represent.

(3) The modules are highly plastic, which allows them to be calibrated and recalibrated by relevant environmental conditions that accumulate over time, or that change, whether naturally or by design of an experimenter (Knudsen, 1988). Thus, stereopsis adjusts at the precognitive level to the changes in binocular disparity that occur as the child's head grows bigger. The phonetic module is similarly calibrated over time according to the phonetic environment to which it is exposed. At all events, the plasticity of these modules is so great that they accommodate stimulus patterns that fall some distance beyond what is possible ecologically. Thus, viewers perceive depth with disparities far greater than could ever be provided by the distance between the eyes. Phonetic perception is possible with a wide variety of departures from the normal acoustic structure of speech, including even sine-wave analogs of the formant tracks.

(4) When the limit of plasticity is exceeded, preemptiveness fails, with the result that heteromorphic and homomorphic representations are evoked simultaneously. Thus, in stereopsis, as the disparity is progressively increased, a point is reached at which the viewer sees heteromorphic depth but also homomorphic disparity. In speech, as the experimenter introduces a discordance or discontinuity between two parts of the signal, a point is reached at which the listener perceives the heteromorphic structure but also the chirps, whistles, or bleats that constitute the homomorphic representation. As it occurs in speech, this phenomenon has come to be known as "duplex perception" (Bentin & Mann, 1990; Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, 1983; Rand, 1974; Whalen & Liberman, 1987). The point to be made here is simply that duplex perception is not a freak phenomenon, limited to speech, but is, rather, what happens to a closed module when, as a consequence of limits on its plasticity, it can no longer preempt the stimulus information.

(5) In the case of stereopsis, it has been shown that, as the disparity is increased over the range of duplex perception, the heteromorphic percept progressively diminishes while the homomorphic percept grows until, finally, only the homomorphic percept is represented (Richards, 1971). (It is as if there were a conservation of stimulus information: some, or all, of the information goes to form the one percept, the remainder goes to the other, and vice versa. Is there, perhaps, some imaginable sense in which the perceptual 'sum' can be said to remain constant?) Mattingly, Yi Xu, and I are currently testing the hypothesis that duplex perception in speech follows a course similar to that found in the duplex range of stereopsis. But whatever the outcome of this test, there is already considerable evidence for the conclusion that the properties of the phonetic module are similar to those that characterize other biological specializations for perception.

In the domain of speech, there are, then, two quite different kinds of biological generality, one for each theory. The horizontal theory claims generality by associating speech with processes that cut across a variety of perceptual, motor, and cognitive functions. The vertical view finds it in the integral relation of speech to language and in the resemblance of speech to other specializations at the precognitive level. The question, then, is not which theory relates speech more generally to other aspects of biology but rather which kind of generality corresponds more closely to the true state of affairs.

The vertical view of speech—that the constituents are gestures, not sounds, and that these constituents are managed by a phonetic specialization—is apparently rejected by most students of speech as implausible and unparsimonious: implausible, because it flies in the face of the common-sense observation that speech consists of sounds that fall on the ear and therefore excite the auditory system; unparsimonious, because it requires the assumption of a distinct and hitherto unacknowledged mode of action and perception. My aim in this paper has been to show

that the shoe is on the other foot. The general form of the argument is that the horizontal view is implausible because the nonlinguistic modalities of action and perception it relies on are manifestly ill suited to the special requirements of phonological communication; it is unparsimonious because it requires cognitive processes of one sort or another if the general auditory and motor units of speech are to be connected to language. The vertical view is designed to avoid these flaws.

REFERENCES

Ades, A. E. (1977). Vowels, consonants, speech and nonspeech. *Psychological Review, 84*, 524-530.

Bentin, S., & Mann, V. A. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88(1)*, 64-74.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of two acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*, 191-211.

Best, C. T., Studdert-Kennedy, M., Manuel, S., & Rubin-Spitz. (1989). Discovering phonetic coherence in auditory patterns. *Perception and Psychophysics, 45*, 237-250.

Bregman, A. (1991). The compositional process in cognition with applications to speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81*, 348-374.

Crowder, R. G., & J. Morton (1969). Pre-categorical acoustic storage (PAS). *Perception & Psychophysics, 5*, 365-373.

DeFrancis, J. (1989). *Visible speech*. Honolulu: University of Hawaii Press.

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121-144.

Diehl, R. L., & Kluender, K. R. (1989). On the categorization of speech sounds. In S. Harnad (Ed.), *Categorical perception*. Cambridge, England: Cambridge University Press.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basic for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America, 85*, 2154-2164.

Fitch, H., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*(4), 343-350.

Fodor, J. (1983) *The modularity of mind*. Cambridge, MA: MIT Press.

Fujisaki, M., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), 29, 207-214.

Howell, P., & Rosen, S. (1984). Natural auditory sensitivities as universal determiners of phonemic contrasts. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for language universals*. The Hague: Mouton.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America, 90*, 83-96.

Kluender, K. R., Diehl, R. L., & Killen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237*, 1195-1197.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Knudsen, E. I. (1988). Experience shapes sound localization and auditory unit properties during development in the barn owl. In G. M. Edelman, W. E. Gall, & M. W. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 137-149). New York: Wiley.

Kuhl, P. K. (1981) Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340-349.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review, 72*, 275-309.

Liberman, A. M. (in press). The relation of speech to reading and writing. Proceedings of the Bellagio Conference on Speech and Reading.

Liberman, A. M., Cooper, F. S., Studdert-Kennedy, M. (1968). Why are spectrograms hard to read? *American Annals of the Deaf, 113*, 127-133.

Liberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception revised. *Cognition, 21*, 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception and Psychophysics*, 30(2), 133-143.

Lindblom, B. (1991). The status of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization. In S. Harnad (Ed.), *Categorical perception*. Cambridge, England: Cambridge University Press.

Mattingly, I. G. (in press). Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning*. Amsterdam: Elsevier Science Publishers.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Functions of the auditory system* (pp. 775-793). New York: Wiley.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2*, 131-157.

Miller, J. D. ((1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. (Life Sciences Research Report 5), p. 49. Berlin: Dahlem Konferenzen.

Oden, G. C., Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172-191.

Parker, E. M. (1988). Auditory constraints on the perception of voice-onset-time: The influence of lower tone frequency on judgments of tone-onset simultaneity. *Journal of the Acoustical Society of America, 83*, 1597-1607.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics, 39*, 129-142.

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception*. Cambridge, England: Cambridge University Press.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13,* 253-260.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics, 34,* 314-322.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society, 55,* 678-680.

Repp, B. H. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. *Psychological Bulletin, 92,* 81-110.

Repp, B. H., & Liberman, A. M. (1987). Phonetic categories are flexible. In S. Harnad (Ed.), *Categorical perception* (pp. 89-112). Cambridge, England: Cambridge University Press.

Richards, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America, 61,* 410-414.

Rosen, S., & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical perception*. Cambridge, England: Cambridge University Press.

Sinnott, J. M. (1976). Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America, 60,* 687-695.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham, (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press.

Stevens, K. N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: Evidence from phonology, acoustics, and psychoacoustics. In T. Myers, J. Laver, & J. Anderson (Eds.), *Advances in psychology: The cognitive representation of speech*. Amsterdam: North Holland.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3-45.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics, 19,* 285-289.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 23,* 169-171.

FOOTNOTE

*Appears in J. Alegria, D. Holender, J. Junça de Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 25-40). Amsterdam: Elsevier Science Publishers B. V. (1992).

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 119-128

The Relation of Speech to Reading and Writing*

Alvin M. Liberman

The difference in naturalness between speech and reading/writing is an important fact for the psychology of language and the obvious point of departure for understanding the processes of literacy, yet it cannot be accounted for by the conventional theory of speech. Because this theory allows no linguistic specialization at the level of perception and action, it necessarily implies that the primary representations of speech are just like those of reading/writing: neither is specifically linguistic, hence both must first be translated into linguistic form if they are to serve a linguistic function. Thus, the effect of the conventional theory is to put speech and reading/writing at the same cognitive remove from language and so make them equally unnatural.

A less conventional view shows the primary motor and perceptual representations of speech to be specifically phonetic, the automatic results of a precognitive specialization for phonological communication. Accordingly, these representations are naturally appropriate for language, requiring no cognitive translation to make them so; in this important respect they differ from the representations of reading/writing. Understanding the source of this difference helps us to see what must be done if readers and writers are to exploit their natural language faculty; why reading and writing should be at least a little difficult for all; and why they might be very difficult for some.

Theories of reading/writing and theories of speech typically have in common that neither takes proper account of an obvious fact about language that must, in any reckoning, be critically relevant to both: there is a vast difference in naturalness (hence ease of use) between its spoken and written forms. In my view, a theory of reading should begin with this fact, but only after a theory of speech has explained it.

My aim, then, is to say how well the difference in naturalness is illuminated by each of two theories of speech—one conventional, the other less so—and then, in that light, to weigh the contribution that each of these can make to an understanding of reading and writing and the difficulties that attend them. More broadly, I aim to promote the notion that a theory of speech and a theory of reading/writing are inseparable, and that the validity of the one is measured, in no small part, by its fit to the other.

This work was supported by NIH grant HD-01994 to Haskins Laboratories. This paper was presented at a conference on "Language and Literacy: Comparative Approaches" in Bellagio, Como, Italy in March, 1991. The conference was organized by Beatrice de Gelder and José Morais.

WHAT DOES IT MEAN TO SAY THAT SPEECH IS MORE NATURAL?

The difference in naturalness between the spoken and written forms of language is patent, so I run the risk of being tedious if I elaborate it here. Still, it is important for the argument I mean to make that we have explicitly in mind how variously the difference manifests itself. Let me, therefore, count the ways.

(1) Speech is universal. Every community of human beings has a fully developed spoken language. Reading and writing, on the other hand, are relatively rare. Many, perhaps most, languages do not even have a written form, and when, as in modern times, a writing system is devised—usually by missionaries—it does not readily come into common use.

(2) Speech is older in the history of our species. Indeed, it is presumably as old as we are, having emerged with us as perhaps the most important of our species-typical characteristics. Writing systems, on the other hand, are developments of the last few thousand years.

(3) Speech is earlier in the history of the individual; reading/writing come later, if at all.

(4) Speech must, of course, be learned, but it need not be taught. For learning to speak, the necessary and sufficient conditions are but two: membership in the human race and exposure to a mother tongue. Indeed, given that these two conditions are met, there is scarcely any way that the development of speech can be prevented. Thus, learning to speak is a precognitive process, much like learning to perceive visual depth and distance or the location of sound. In contrast, reading and writing require to be taught, though, given the right ability, motivation, and opportunity, some will infer the relation of script to language and thus teach themselves. But, however learned, reading/writing is an intellectual achievement in a way that learning to speak is not.

(5) There are brain mechanisms that evolved with language and that are, accordingly, largely dedicated to its processes. Reading and writing presumably engage at least some of these mechanisms, but they must also exploit others that evolved to serve nonlinguistic functions. There is no specialization for reading/writing as such.

(6) Spoken language has the critically important property of 'openness': unlike nonhuman systems of communication, speech is capable of expressing and conveying an indefinitely numerous variety of messages. A script can share this property, but only to the extent that it somehow transcribes its spoken-language base. Having no independent existence, a proper (open) script is narrowly constrained by the nature of its spoken-language roots and by the mental resources on which they draw. Still, within these constraints, scripts are more variable than speech.

One dimension of variation is the level at which the message is represented, though the range of that variation is, in fact, much narrower than the variety of possible written forms would suggest. Thus, as DeFrancis (1989) convincingly argues, any script that communicates meanings or ideas directly, as in ideograms, for example, is doomed to arrive at a dead end. Ideographic scripts cannot be open—that is, they cannot generate novel messages—and the number of messages they can convey is never more than the inventory of one-to-one associations between (holistically different) signals and distinctly different meanings that human beings can master. Indeed, it is a distinguishing characteristic of language, and a necessary condition of its openness, that it communicates meanings indirectly, via specifically linguistic structures and processes, including, nontrivially, those of the phonological component. Not surprisingly, scripts must follow suit; in the matter of language, as with so many other natural processes, it is hard to improve on nature.

Constraints of a different kind apply at the lower levels. Thus, the acoustic signal, as represented visually by a spectrogram, for example, cannot serve as a basis for a script; while spectrograms can be puzzled out by experts, they, along with other visual representations, cannot be read fluently. The reason is not primarily that the relevant parts of the signal are insufficiently visible; it is, rather, that, owing to the nature of speech, and especially to the coarticulation that is central to it, the relation between acoustic signal and message is complex in ways that defeat whatever cognitive processes the 'reader' brings to bear. Narrow phonetic transcriptions are easier to read, but there is still more context-, rate-, and speaker-conditioned variation than the eye is comfortable with. In any case, no extant script offers language at a narrow phonetic level. To be usable, scripts must, apparently, be pitched at the more abstract phonological and morphophonological levels. That being so, and given that reading-writing require conscious awareness of the units represented by the script, we can infer that people can become conscious of phonemes and morphophonemes. We can also infer about these units that, standing above so much of the acoustic and phonetic variability, they correspond approximately to the invariant forms in which words are presumably stored in the speaker's lexicon. A script that captures this invariance is surely off to a good start. At all events, some scripts (e.g., Finnish, Serbo-Croatian) do approximate to purely phonological renditions of the language, while others depart from a phonological base in the direction of morphology. Thus, English script is rather highly morphophonological, Chinese even more so. But, as DeFrancis (1989; see also Wang, 1981) makes abundantly clear, all these scripts, including even the Chinese, are significantly phonological, and, in his view, they would fail if they were not; the variation is simply in the degree to which some of the morphology is also represented.

Scripts also vary somewhat, as speech does not, in the size of the linguistic segments they take as their elements, but here, too, the choice is quite constrained. Surely, it would not do to make a unit of the script equal to a phoneme and a half, a third of a syllable, or some arbitrary stretch—say 100 milliseconds—of the speech stream. Still, scripts can and do take as their irreducible units either phonemes or syllables, so in this respect, too, they are more diverse than speech.

(7) All of the foregoing differences are, of course, merely reflections of one underlying circumstance—namely, that speech is a product of biological evolution, while writing systems are artifacts. Indeed, an alphabet—the writing system that is of most immediate concern to us—is a triumph of applied biology, part discovery, part invention. The discovery—surely one of the most momemtous of all time—was that words do not differ from each other holistically, but rather by the particular arrangement of a small inventory of the meaningless units they comprise. The invention was simply the notion that if each of these units were to be represented by a distinctive optical shape, then everyone could read and write, provided he knew the language and was conscious of the internal phonological structure of its words.

HOW IS THE DIFFERENCE IN NATURALNESS TO BE UNDERSTOOD?

Having seen in how far speech is more natural than reading/writing, we should look first for a simple explanation, one that is to be seen in the surface appearance of the two processes. But when we search there, we are led to conclude, in defiance of the most obvious facts, that the advantage must lie with reading/writing, not with speech. Thus, it is the eye, not the ear, that is the better receptor; the hand, not the tongue, that is the more versatile effector; the print, not the sound, that offers the better signal-to-noise ratio; and the discrete alphabetic characters, not the nearly continuous and elaborately context-conditioned acoustic signal, that offers the more straightforward relation to the language. To resolve this seeming paradox and understand the issue more clearly, we shall have to look more deeply into the biology of speech. To that end, I turn to two views of speech to see what each has to offer.

The conventional view of speech as a basis for understanding the difference in naturalness. The first assumption of the conventional view is so much taken for granted that it is rarely made explicit. It is, very simply, that the phonetic elements are defined as sounds. This is not merely to say the obvious, which is that speech is conveyed by an acoustic medium, but rather to suppose, in a phrase made famous by Marshall McLuhan, that the medium *is* the message.

The second assumption, which concerns the production of these sounds, is also usually unspoken, not just because it is taken for granted, though it surely is, but also because it is apparently not thought by conventional theorists to be even relevant. But, whatever the reason, one finds among the conventional claims none which implies the *existence* of a phonetic mode of action—that is, a mode adapted to phonetic purposes and no other. One therefore infers that the conventional view must hold (by default, as it were) that no such mode exists. Put affirmatively, the conventional assumption is that speech is produced by motor processes and movements that are independent of language.

The third assumption concerns the perception of speech sounds, and, unlike the first two, is made explicitly and at great length (Cole & Scott, 1974; Crowder & Morton, 1969; Diehl & Kluender, 1989; Fujisaki & Kawashima, 1970; Kuhl, 1981; Miller, 1977; Oden & Massaro, 1978; Stevens, 1975). In its simplest form, it is that perception of speech is not different from perception of other sounds; all are governed by the same general processes of the auditory system. Thus, language simply accepts representations made available to it by perceptual processes that are generally auditory, not specifically linguistic. So, just as language presumably recruits ordinary motor processes for its own purposes, so, too, does it recruit the ordinary processes of auditory perception; at the level of perception, as well as action, there is, on the conventional view, no specialization for language.

The fourth assumption is required by the second and third. For if the acts and percepts of speech are not, by their nature, specifically phonetic, they must necessarily be made so, and that can be done only by a process of cognitive translation. Presumably, that is why conventional theorists say about speech perception that after the listener has apprehended the auditory representation he must elevate it to linguistic status by attaching a phonetic label (Crowder & Morton, 1969; Fujisaki & Kawashima, 1970; Pisoni, 1973), fitting it to a phonetic prototype (Massaro, 1987; Oden & Massaro, 1978), or associating it with some other

linguistically significant entity, such as a 'distinctive feature' (Stevens, 1975).

I note, parenthetically, that this conventional way of thinking about speech is heir to two related traditions in the psychology of perception. One, which traces its origins to Aristotle's enumeration of the five senses, requires of a perceptual mode that it have an end organ specifically devoted to its interests. Thus, ears yield an auditory mode; eyes, a visual mode; the nose, an olfactory mode; and so on. Lacking an end organ of its very own, speech cannot, therefore, be a mode. In that case, phonetic percepts cannot be the immediate objects of perception; they can only be perceived secondarily, as the result of a cognitive association between a primary auditory representation appropriate to the acoustic stimulus that excites the ear (and hence the auditory mode) and, on the other hand, some cognitive form of a linguistic unit. Such an assumption is, of course, perfectly consistent with another tradition in psychology, one that goes back at least to the beginning of the 18th century, where it is claimed in Berkeley's "New Theory of Vision" (1709) that depth (which cannot be projected directly onto a two-dimensional retina) is perceived by associating sensations of muscular strain (caused by the convergence of the eyes as they fixate objects at various distances) with the experience of distance. In the conventional view of speech, as in Berkeley's assumption about visual depth, apprehending the event or property is a matter of perceiving one thing and calling it something else.

Some of my colleagues and I have long argued that the conventional assumptions fail to account for the important facts about speech. Here, however, my concern is only with the extent to which they enlighten us about the relation of spoken language to its written derivative. That the conventional view enlightens us not at all becomes apparent when one sees that, in contradiction of all the differences I earlier enumerated, it leads to the conclusion that speech and reading/writing must be equally natural. To see how comfortably the conventional view sits with an (erroneous) assumption that speech and reading/writing are psychologically equivalent, one need only reconsider the four assumptions of that view, substituting, where appropriate, 'optical' for 'acoustic' or 'visual' for 'auditory.'

One sees then, that, just as the phonetic elements of speech are, by the first of the conventional assumptions, defined as sounds, the elements of a writing system can only be defined as optical shapes. As for the second assumption—viz., that speech production is managed by motor processes of the most general sort—we must suppose that this is exactly true for writing; by no stretch of the imagination can it be supposed that the writer's movements are the output of an action mode that is specifically linguistic. The third assumption of the conventional view of speech also finds its parallel in reading/writing, for, surely, the percepts evoked by the optical characters are ordinarily visual in the same way that the percepts evoked by the sounds of speech are supposed to be ordinarily auditory. Thus, at the level of action and perception, there is in reading/writing, as there is assumed to be in speech, no specifically linguistic mode. For speech, that is only an assumption—and, as I think, a very wrong one—but for reading/writing it is an incontrovertible fact; the acts and percepts of reading/writing did not evolve as part of the specialization for language, hence they cannot belong to a natural linguistic mode.

The consequence of all this is that the fourth of the conventional assumptions about speech is, in fact, necessary for reading/writing and applies perfectly to it: like the ordinary, nonlinguistic auditory and motor representations according to conventional view of speech, the correspondingly ordinary visual and motor representations of reading/writing must somehow be made relevant to language, and that can only be done by a cognitive process; the reader/writer simply has to learn that certain shapes refer to units of the language and that others do not.

It is this last assumption that most clearly reveals the flaw that makes the conventional view useless as a basis for understanding the most important difference between speech and reading/writing—namely, that the evolution of the one is biological, the other cultural. To appreciate the nature of this shortcoming, we must first consider how either mode of language transmission meets a requirement that is imposed on every communication system, whatever its nature and the course of its development. This requirement, which is commonly ignored in arguments about the nature of speech, is that the parties to the message exchange must be bound by a common understanding about which signals, or which aspects of which signals, have communicative significance; only then can communication succeed. Mattingly and I have called this the requirement for 'parity' (Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Mattingly & Liberman, 1988). One asks, then, what is entailed by parity as the system develops in the species

and as it is realized in the normal communicative act.

In the development of writing systems, the answer is simple and beyond dispute: parity was established by agreement. Thus, all who use an alphabet are parties to a compact that prescribes just which optical shapes are to be taken as symbols for which phonological units, the association of the one with the other having been determined arbitrarily. Indeed, this is what it means to say that writing systems are artifacts, and that the child's learning the linguistic significance of the characters of the script is a cognitive activity.

Unfortunately for the validity of the conventional assumptions, they require that the same story be told about the development of parity in speech. For if the acts and percepts of speech are, as the conventional assumption would have it, ordinarily motor and ordinarily auditory, one must ask how, why, when, and by whom they were invested with lingistic significance. Where is it written that the gesture and percept we know as [b] should count for language, but that a clapping of the hands should not? Is there somewhere a commandment that says, Thou shalt not commit [b] except when it is thy clear intention to communicate? Or are we to assume, just as absurdly, that [b] was incorporated into the language by agreement? It is hard to see how the conventional view of speech can be made to provide a basis for understanding the all-important difference in evolutionary status between speech and reading/writing.

The problem is the worse confounded when we take account of both sides of the normal communicative act. For, on the conventional view the speaker deals in representations of a generally motor sort and the listener in representations of a generally auditory sort. What is it, then, that these two representations have in common, except that neither has anything to do with language? One must thus suppose for speech, as for writing and reading, that there is something like a phonetic idea—a cognitive representation of some kind—to connect these representations to each other and to language, and so to make communication possible.

Thus it is that at every biological or psychological turn the conventional view of speech make reading and writing the equivalents of speech perception and production. Since these processes are plainly not equivalent, the conventional view of speech can hardly be the starting point for an account of reading and writing.

The unconventional view of speech as a basis for understanding the difference in naturalness. The first assumption of the unconventional view is that the units of speech are defined as gestures, not as the sounds that those gestures produce. (For recent accounts of the unconventional view, see: Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Mattingly & Liberman, 1988; Mattingly & Liberman, 1990). The rationale for this assumption is to be understood by taking account of the function of the phonological component of the grammar and of the requirements it imposes. As for the function of phonology, it is, of course, to form words by combining and permuting a few dozen meaningless segments, and so to make possible a lexicon tens of thousands of times larger than could ever have been achieved if, as in all natural but nonhuman communication systems, each 'word' were conveyed by a signal that was holistically different from all others. But phonology can serve this critically important function only if its elements are commutable; and if they are to be commutable, they must be discrete and invariant.

A related requirement has to do with rate, for if all utterances are to be formed by variously stringing together an exiguous set of signal elements, then, inevitably, the strings must run to great lengths. It is essential, therefore, if these strings are to be organized into words and sentences, that they be produced and perceived at reasonable speed. But if the auditory percepts of the conventional view are to be discrete and invariant, the sounds and gestures must be discrete and invariant, too. Such sounds and gestures are possible, of course, but only at the expense of rate. Thus one could not, on the conventional view, say 'bag,' but only [b] [a] [g], and to say [b] [a] [g] is not to speak but to spell. Of course, if speech were like that, then everyone who could speak or perceive a word would know exactly how to write and read it, provided only that he had managed the trivial task of memorizing the letter-to-sound correspondences. The problem is that there would be no language worth writing or reading.

There seems, indeed, no way to solve the rate problem and still somehow preserve the acoustic-auditory strategy of the conventional view. It would not have helped, for example, if Nature had abandoned the vocal tract and equipped her

human creatures with acoustic devices adapted to producing a rapid sequence of sounds—a drumfire or tattoo—for that strategy would have defeated the ear. The point is that speech proceeds at rates that transmit up to 15 or even 20 phonemes per second, but if each phoneme were represented by a discrete sound, then rates that high would seriously strain and sometimes overreach the ability of the ear to resolve the individual sounds and to divine their order.

According to the unconventional view, Nature solved the problem by avoiding the acoustic-auditory strategy that would have created it. The alternative she chose was to define the phonetic elements as gestures, as the first assumption of the unconventional view proposes. Thus, [b] is a closing at the lips, [h] an opening at the glottis, [p] a combination of lip closing and glottis opening, and so forth. In fact, the gestures are far more complex than this, for a gesture usually comprises movements of several articulators, and these movements are exquisitely context-conditioned. Given such complications, I must wait on others to discover how best to characterize these gestures and how to derive the articulatory movements from them. But while I'm waiting, I can be reasonably sure that the unconventional view heads the theoretical enterprise in the right direction, for it permits coarticulation. That is, it permits the speaker to overlap gestures that are produced by different organs—for example, the lips and the tongue in [ba]—and to merge gestures that are produced by different parts of the same organ—for example, the tip and body of the tongue, as in [da]—and so to achieve the high rates that are common.

But the gestures that are coarticulated, and the means for controlling them, were not lying conveniently to hand, just waiting to be appropriated by language, which brings us to the second assumption of the unconventional view: the gestures of speech and their controls are specifically phonetic, having been adapted for language and for nothing else. As for the gestures themselves, they are distinct as a class from those movements of the same organs that are used for such nonlinguistic purposes as swallowing, moving food around in the mouth, licking the lips, and so on. Presumably, they were selected in the evolution of speech in large part because of the ease with which they lent themselves to being coarticulated. But the control and coordination of these gestures is specific to speech, too. For coarticulation must walk a fine line, being constrained on either side by the special demands of phonological communication. Thus, coarticulation must produce enough overlap and merging to permit the high rates of phonetic segment production that do, in fact, occur, while yet preserving the details of phonetic structure.

The third assumption of the unconventional view is that, just as there is a specialization for the production of phonetic structures, so, too, is there a specialization for their perception. Indeed, the two are but complementary aspects of the same specialization, one for deriving the articulatory movements from the (abstract) specification of the gestures, the other for processing the acoustic signals so as to recover the coarticulated gestures that are its distal cause. The rationale for this assumption about perception arises out of the consequences of the fact that coarticulation folds information about several gestures into a single piece of sound, thereby conveying the information in parallel. This is of critical importance for language because it relaxes by a large factor the constraint on rate of phonetic-segment perception that is set by the temporal resolving power of the ear. But this gain has a price, for coarticulation produces a complex and singularly linguistic relation between acoustic signal and the phonetic message it conveys. As is well known, the signal for each particular phonetic element is vastly different in different contexts, and there is no direct correspondence in segmentation between signal and phonetic structure. It is to manage this language-specific relation between signal and appropriate percept that the specialization for speech perception is adapted. Support for the hypothesis that there is such a specialized speech mode of perception is to be found elsewhere. (See references given at the beginning of this section.) What is important for our present purposes is only that, according to this hypothesis, the percepts evoked by the sounds of speech are immediately and specifically phonetic. There is no need, as there is on the conventional view, for a cognitive translation from an initial auditory representation, simply because there is no initial auditory representation.

Now one can see plainly the difference between speech and reading/writing. In reading, to take the one case, the primary perceptual representations are, as we have seen, inherently visual, not linguistic. Thus, these representations are, at best, arbitrary symbols for the natural units of language, hence unsuited to any natural language process until and unless they have been

translated into linguistic form. On the other hand, the representations that are evoked by the sounds of speech are immediately linguistic in kind, having been made so by the automatic processes of the phonetic module. Accordingly, they are, by their very nature, perfectly suited for the further automatic and natural processing that the larger specialization for language provides.

As for parity and its development in evolution and in the child, it is, on the unconventional view, built into the very bones of the system. For what evolved, on this view, was a specifically phonetic process, together with representations that were thus categorically set apart from all others and reserved for language. The unconventional view also allows us to see, as the link between sender and receiver, the specifically phonetic gestures that serve as the common coin for the conduct of their linguistic business. There is no need to establish parity by means of (innate) phonetic ideas—e.g., labels, prototypes, distinctive features—to which the several nonlinguistic representations must be cognitively associated.

HOW CAN READING/WRITING BE MADE TO EXPLOIT THE MORE NATURAL PROCESSES OF SPEECH?

The conventional view of speech provides no basis for asking this question, since there exists, on this view, no difference in naturalness. It is perhaps for this reason that the (probably) most widely held theory of reading in the United States explicitly takes as its premise that reading and writing are, or at least can be, as natural and easy as speech (Goodman & Goodman, 1979). According to this theory, called 'whole language,' reading and writing prove to be difficult only because teachers burden children with what the theorists call "bite-size abstract chunks of language such as words, syllables, and phonemes" (Goodman, 1986). If teachers were to teach children to read and write the way they were (presumably) taught to speak, then there would be no problem. Other theorists simply ignore the primacy of speech as they describe a reading process in which purely visual representations are sufficient to take the reader from print to meaning, thus implying a 'visual' language that is somehow parallel to a language best described as 'auditory' (see, for example, Massaro & Schmuller, 1975; F. Smith, 1971).

On the unconventional view, however, language is neither auditory nor visual. If it seems to be auditory, that is only because the appropriate stimulus is commonly acoustic (*pace* Aristotle). But optical stimuli will, under some conditions, evoke equally convincing phonetic percepts, provided (and this is a critical proviso) they specify the same articulatory movements (hence, phonetic gestures) that the sounds of speech evoke. This so-called 'McGurk effect' works powerfully when the stimuli are the natural movements of the articulatory apparatus, but not when they are the arbitrary letters of the alphabet. Thus, language is a mode, largely independent of end organs, that comprises structures and processes specifically adapted to language, hence easy to use for linguistic purposes. Therefore, the seemingly sensible strategy for the reader is to get into that mode, for once there, he is home free; everything else that needs to be done by way of linguistic processing is done for him automatically by virtue of his natural language capacity. As for where the reader should enter the language mode, one supposes that earlier is better, and that the phonological component of the mode is early enough. Certainly, making contact with the phonology has several important advantages: it makes available to the reader a generative scheme that comprehends all the words of the language, those that died yesterday, those that live today, and those that will be born tomorrow; it also establishes clear and stable representations in a semantic world full of vague and labile meanings; and, not least, it provides the natural grist for the syntactic mill—that is, the phonological representations that are used by the working memory as it organizes words into sentences.

The thoroughly visual way to read, described earlier, is the obvious alternative, doing everything that natural language does without ever touching its structures and processes. But surely that must be a hard way to read, if, indeed, it is even possible, since it requires the reader to invent new and cognitively taxing processes just in order to deal with representations that are not specialized for language and for which he has no natural bent.

WHAT OBSTACLE BLOCKS THE NATURAL PATH?

As we have seen, the conventional view allows two equivalent representations of language—one auditory, the other visual—hence two equally natural paths that language processes might follow. In that case, such obstacles as there might

be could be no greater for the visual mode; indeed, accepting the considerations I mentioned earlier, we should have to suppose that visual representations would offer the easier route.

The unconventional view, on the other hand, permits one to see just what it is that the would-be reader and writer (but not the speaker/listener) must learn, and why the learning might be at least a little difficult. The point is that, given the specialization for speech, anyone who wants to speak a word is not required to know how it is spelled; indeed, he does not even have to know that it has a spelling. He has only to think of the word; the speech specialization spells it for him, automatically selecting and coordinating the appropriate gestures. In an analogous way, the listener need not consciously parse the sound so as to identify its constituent phonological elements. Again, he relies on the phonetic specialization to do all the hard work; he has only to listen. Because the speech specialization is a module, its processes are automatic and insulated from consciousness. There are, therefore, no cognitively formed associations that would make one aware of the units being associated. Of course, the phonological representations, as distinguished from the processes, are not so insulated; they are available to consciousness—indeed, if they were not, alphabetic scripts would not work—but there is nothing in the ordinary use of language that requires the speaker/listener to put his attention on them. The consequence is that experience with speech is normally not sufficient to make one consciously aware of the phonological structure of its words, yet it is exactly this awareness that is required of all who would enjoy the advantages of an alphabetic scheme for reading and writing.

Developing an awareness of phonological structure, and hence an understanding of the alphabetic principle, is made the more difficult by the coarticulation that is central to the function of the phonetic specialization. Though such coarticulation has the crucial advantage of allowing speech production and perception to proceed at reasonable rates, it has the disadvantage from the would-be reader/writer's point of view that it destroys any simple correspondence between the acoustic segments and the phonological segments they convey. Thus, in a word like 'bag,' coarticulation folds three phonological segments into one seamless stretch of sound in which information about the several phonological segments is thoroughly overlapped. Accordingly, it avails the reader little to be able to identify the letters, or even to know their sounds. What he must know, if the script is to make sense, is that a word like 'bag' has three pieces of phonology even though it has only one piece of sound. There is now much evidence (1) that preliterate and illiterate people (large and small) lack such phonological awareness; (2) that the amount of awareness they do have predicts their success in learning to read, and (3) that teaching phonological awareness makes success in reading more likely. (For a summary, see, for example, I. Y. Liberman & A. M. Liberman, 1990).

WHY SHOULD THE OBSTACLE LOOM ESPECIALLY LARGE FOR SOME?

Taking the conventional view of speech seriously makes it hard to avoid the assumption that the trouble with the dyslexic must be in the visual system. It is, therefore, not in the least surprising to find that by far the largest number of theories about dyslexia do, in fact, put the problem there. Thus, some believe that the trouble with dyslexics is that they cannot control their eye movements (Pavlides, 1981), or that they have problems with vergence (Stein, Riddell, & Fowler, 1989) or that they see letters upside down or wrong side to (Orton, 1937), or that their peripheral vision is better than it should be (Geiger & Lettvin, 1989), and so on.

The unconventional view of speech directs one's attention, not to the visual system and the various problems that might afflict it, but rather to the specialization for language and the reasons why the alphabetic principle is not self-evident. As we have seen, this view suggests that phonological awareness, which is necessary for application of the alphabetic principle, does not come for free with mastery of the language. As for dyslexics—that is, those who find it particularly hard to achieve that awareness—the unconventional view of speech suggests that the problem might well arise out of a malfunction of the phonological specialization, a malfunction sufficient to cause the phonological representations to be less robust than normal. Such representations would presumably be just that much harder to become aware of. While it is difficult to test that hypothesis directly, it is possible to look for support in the other consequences that a weak phonological faculty should have. Thus, one would expect that dyslexics would show such other symptoms as greater-than-normal difficulty in holding and manipulating verbal (but not

nonverbal) materials in working memory, in naming objects (that is, in finding the proper phonological representation), in perceiving speech (but not nonspeech) in noise, and in managing difficult articulations. There is some evidence that dyslexics do show such symptoms. (For a summary, see: I. Liberman, Shankweiler, & A. Liberman, 1985).

WHAT ARE THE IMPLICATIONS FOR A THEORY OF SPEECH?

Those who investigate the perception and production of speech have been little concerned to explain how these processes differ so fundamentally in naturalness from those of reading and writing. Perhaps this is because the difference is so obvious as to be taken for granted and so to escape scientific examination. Or perhaps the speech researchers believe that explaining the difference is the business of those who study reading and writing. In any case, neglect of the difference might be justifiable if it were possible for a theory of speech to have no relevant implications. But a theory of speech does inevitably have such implications, and, as has been shown, the implications of the conventional theory run counter to the obvious facts. My concern in this paper has been to show that, as a consequence, the conventional theory is of little help to those who would understand reading and writing. Now I would suggest that, for exactly the same reason, the theory offers little help to those who would understand speech, for if the theory fails to offer a reasonable account of a most fundamental fact about language, then we should conclude that there is something profoundly wrong with it.

The unconventional theory of speech described in this paper was developed to account for speech, not for the difference between its processes and those of reading and writing. That it nevertheless shows promise of also serving the latter purpose may well be taken as one more reason for believing it.

REFERENCES

Berkeley, G. (1709). *An essay towards a new theory of vision.* Dublin: Printed by Aaron Rhames for Jeremy Pepyal.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81,* 348-374.

Crowder, R. G., & Morton J. (1969) Pre-categorical acoustic storage (PAS). *Perception & Psychophysics, 5,* 365-373.

DeFrancis, J. (1991). *Visible speech: The diverse oneness of writing systems.* Honolulu: University of Hawaii Press.

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.

Fujisaki, M., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), *29,* 207-214.

Geiger, G., & Lettvin, J. Y. (1988). Dyslexia and reading as examples of alternative visual strategies. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading.* London: The Macmillan Press Ltd.

Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, DE: International Reading Association.

Goodman, K. S., & Goodman, Y. M. (1979). Learning to read is natural. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading,* Vol. 1. Hillsdale, NJ: Erlbaum.

Goodman, K. S. (1986). *What's whole in whole language: A parent-teacher guide.* Portsmouth, NH: Heinemann.

Kuhl, P. K. (1981) Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70,* 340-349.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489-494.

Liberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Annals of Dyslexia, 40.*

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle.* Research Monograph Series. Ann Arbor: University of Michigan Press.

Mattingly, I. G., & A. M. Liberman. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Functions of the auditory system* (pp. 775-793). New York: Wiley.

Mattingly, I. G. & A. M. Liberman. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps.* New York: Wiley.

Miller, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals.* (Life Sciences Research Report 5). Berlin: Dahlem Konferenzen.

Oden, G. C., & D. W. Massaro (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172-191.

Orton, S. J. (1937). *Reading, writing and speech problems in children.* New York: W. W. Norton & Co.

Pavlides, G. T. (1985). Eye movement differences between dyslexics, normal, and retarded readers while sequentially fixating digits. *American Journal of Opto and Physiological Optics, 62,* 820-832.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13,* 253-260.

Stein, J., Riddell, P., & Fowler, S. (1988). Disordered right hemisphere function in developmental dyslexia. In C. von

Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Brain and Reading*. England: The Macmillan Press Ltd.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press.

Wang, W. S-Y. (1981). Language structure and optimal orthography. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

FOOTNOTE

*A slightly different version of this paper appears in R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 167-178). Amsterdam: Elsevier Science Publishers (1992).

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 129-140

Linguistic Awareness and Orthographic Form*

Ignatius G. Mattingly†

INTRODUCTION: THE TAXONOMY OF WRITING SYSTEMS

To impose some pattern on the vast array of writing systems, present and past,[1] several investigators have proposed typologies of writing (Gelb, 1963; Hill, 1967; Sampson, 1985; DeFrancis, 1989; see DeFrancis for a review). While typology for its own sake may seem a dubious goal, these proposals bring to notice certain interesting questions.

Consider first the problem posed by logograms. It is generally recognized that the signs found in writing fall into two broad categories: logographic and phonographic. Logograms stand for words, or more precisely, morphemes. Thus, in Sumerian writing, there is a logogram that stands for the morpheme **ti**, 'arrow.' Phonographic signs stand for something phonological: syllables or phonemic segments. Thus, in Old Persian, there is a sign for the syllable **da**, and in Greek alphabetic writing, a sign for the vowel **a**. This distinction suggests that writing systems might be classified according to whether they are logographic or phonographic. But the attempt to impose such a classification is embarrassed by the fact that while the many systems in the West Semitic tradition are indeed essentially phonographic and have no logograms, writing systems of all other traditions use both logograms and phonograms. There have been no purely logographic systems: phonographic signs are found in all traditions.

In these circumstances, Gelb sets up a hybrid category "word-syllabic," in which he includes Sumerian, Egyptian (whose phonographic signs he takes to be syllabic[2]), and Chinese. Other orthographic taxonomists allow a writing system to belong to two different categories. Thus for Hill, Egyptian is both "phonemic" and "morphemic" and for Sampson, Japanese is both "phonographic" and "logographic." DeFrancis, recognizing that logograms are neither necessary nor sufficient for an orthography, more sensibly treats logography as an optional accompaniment to various phonographic categories. But the question of interest is *why* logograms should play only this secondary role, why there have been no pure logographies.

A second problem arises in sorting out the phonographic categories. Here one might recognize, with DeFrancis, systems like Sumerian or Linear B, in which the phonographic signs stands for syllables; systems like Egyptian or Phoenician, in which they stand for consonants; and systems like Greek or English, in which they stand for both consonants and vowels (*plene* systems).

The distinction between consonantal and *plene* systems, however, proves to be less than rigid. In Egyptian, the letters for **j**, **w**, and **ʔ** are used to write **i**, **u** and **a**, respectively, in foreign names (Gelb, 1963). Phoenician, indeed, is a strictly consonantal, but the other "consonantal" systems deriving from it all have some convention for transcribing vowels when necessary. For example, in Aramaic, the letters *yodh*, *waw*, and *he* (or *aleph*) were used to write final **i**, **u**, and **a**, respectively, and to render vowels in foreign names (Cross & Freedman, 1952). In Masoretic Hebrew, Arabic, and various Indic systems, vowels are regularly indicated by diacritic marks on consonant letters. And, of course, the first clearly *plene* system, the Greek alphabet, is a development from the Phoenician consonantal system. The taxonomist thus has to decide where to draw the line between essentially consonantal systems, hybrid systems, and undoubted *plene* systems. Perhaps the wisest course is the one followed by Sampson: simply to classify all these systems as "segmental."

Preparation of this paper was supported in part by NICHD grant HD-01994 to Haskins Laboratories. Alice Faber, Leonard Katz, Alvin Liberman, and Yi Xu gave helpful comment and criticism on an earlier draft.

Syllabic systems, in contrast, are clearly a separate category and present no problem to the taxonomist. There is no writing system that must be regarded as a hybrid between a syllabic and a segmental system. Syllabic systems show no tendency to analyze syllables into segments. What is found, rather, is that when analysis becomes necessary, complex syllables are analyzed into simpler syllables. Thus, neither the Mesopotamian nor the Mayan syllabaries had signs for all possible $C_1V_1C_2$ syllables in their respective languages. Instead, such syllables were written in Mesopotamian as if they were C_1V_1 + V_1C_2 (Driver, 1976) and in Mayan as if they were C_1V_1 + C_2V_1 (Kelley, 1976)). Similarly, Greek $C_1C_2V_1$...syllables were written in Linear B as C_1V_1+ C_2V_1 +...(Ventris & Chadwick, 1973). Nor, despite suggestions to the contrary by Gelb and DeFrancis, has a syllabic system ever developed into a segmental system, or conversely.[3] It cannot be excluded that the Egyptians may, as DeFrancis says (following Ray, 1986), have gotten the idea of writing from the Sumerians. But there is certainly no reason to believe that they borrowed the idea of *syllabic* writing from the Sumerians and then adapted it to consonantal writing, in the way that the Greeks may be said to have borrowed the idea of *consonantal* writing from the Phoenicians and adapted it to *plene* writing. The various orthographic traditions are remarkably self-consistent in this matter. The Mesopotamian, Chinese, Cretan and Mayan traditions began and remained syllabic; the Egyptian and West Semitic traditions began and remained segmental.

If the main purpose here were to arrive at a taxonomy of writing systems, the conclusion would have to be that there are two primary categories: syllabic and segmental. Either of these may or may not be accompanied by logograms. Transcription of vowels in segmental systems is a matter of degree, with Phoenician at one end of the scale and Greek at the other. The interesting question, however, particularly given the degree of overlap or hybridization that is found between logographic and phonographic categories, and between consonantal and *plene* categories, is why the syllabic and segmental categories have remained so distinct.

In an attempt to answer the questions just posed, it is necessary to consider why an orthography can make reading and writing possible, what constraints there are on the form of orthographies, how orthographies could have been invented, and what happens when orthographies are transmitted from one culture to another.

WHY READING AND WRITING ARE POSSIBLE[4]

When a listener has just heard an utterance in a language he knows, he has available for a brief time not only his understanding of the semantic and pragmatic content of the utterance (the speaker's *message*), but also a mental representation of its linguistic structure. The basis for this claim is that a linguist, by analyzing the intuitions of informants about utterances in their native language (such as that two utterances are or are not the same word, or that a certain word is the subject of a sentence), can formulate a coherent grammar, consistent with grammars that would be formulated by other linguists working with other informants on the same language. This holds true even if, as is typically the case for a language with no writing system, the informants are quite unaware of the linguistic units into which utterances in their language can be analyzed. Because the informants' intuitions are apparently valid, they must be based on linguistic representations of some kind.

While linguists are not in total agreement about the nature of the linguistic representation of an utterance, it seems reasonably clear that such a representation must include the syntactic structure, the selection of lexical items and their component morphemes, the phonological structure, and the phonetic structure. The linguist's syntactic diagrams and phonological and phonetic transcriptions are formal reconstructions of different levels of the representation. These levels are not independent of one another. Syntax constrains lexical choice, lexical choice determines morphology and phonology, syntax and phonology determine phonetic structure. The representation thus has extensive inherent redundancy.

The linguistic representation is strictly structural rather than procedural. The listener has no access to the many intermediate steps he must presumably go through in the course of parsing the utterance, so that these steps are not represented. Acoustic details such as formant trajectories are not part of the linguistic representation, simply because the listener does not perceive them as such, but only the phonetic events they reflect. Other aspects of the utterance, such as individual voice quality, speaking rate, and loudness, which the listener can hear, must be presumed to be excluded because they are not linguistic at all and never serve to mark a linguistic difference between two utterances.

Access must be distinguished from awareness. All normal language users, it has been claimed, have access to the contents of linguistic representations. This means that they have a potential ability to introspect and report on significant details of the representation, and to regard it as a structure of phrases, words, and segments, not that they can actually do so. The representation is a complicated affair, and a person who is not "linguistically aware" can no more be expected to notice its characteristic units and structure than an electronically naive person can be expected to appreciate the units and structure of a circuit diagram (Mattingly, 1972). Linguistic awareness must in large part be acquired. The principal stimulus for linguistic awareness in modern cultures is literacy (Morais, Cary, Alegria, & Bertelson, 1979). Unlike illiterate adults or preliterate children, those who have learned to read can readily report on and manipulate at least those units of the linguistic representations of spoken utterances to which units of the orthography correspond (Read, Zhang, Nie, & Ding, 1986). However, there must certainly be other sources of linguistic awareness: Long before writing was known, poets composed verse in meters requiring strict attention to subtle phonological details.

It is not agreed how linguistic representations are created. On one view, they are a byproduct of the cognitive processes by which utterances are analyzed. Linguistic information, recovered step by step from the auditory image of the input signal, is temporarily represented in memory until, at a later stage, the speaker's message can be computed (Baddeley, 1986). The difficulty with this view is that, as has been noted, the language user seems to have no access to the supposedly cognitive analytic steps that must precede the formation of the representation or to the subsequent steps by which the message is derived from this representation. An alternative view is that the representation, as well as the message itself, is not a byproduct but a true output of a specialized, low-level processor (the "language module") whose internal operations, being inaccessible to cognition, have no cognitive byproducts (Fodor, 1983). This view implies that the linguistic representation must have some biological function other than communication, for which the message alone would suffice. What this function might be is unclear (but see Mattingly, 1991, for some speculations).

So far, the cognitive linguistic representation has been considered just as the product of the perception of utterances. But such representations are produced in the course of other modes of linguistic processing as well. Thus, a linguistic representation is formed in the production of an utterance, so that the speaker knows what it is he has just said. And when one rehearses an utterance in order to keep it in mind verbatim, what presumably happens is that the linguistic processor uses a decaying linguistic representation to construct a fresh version of the representation, and incidentally, of the message. This seeming defiance of entropy is possible for linguistic representations (as it may not be for mental representations in general) because of their high inherent redundancy.

Consideration of rehearsal also shows that the linguistic representation can be an input to as well as an output from the linguistic processor. Even more significantly, for the present purposes, a representation not originally produced by primary processes of perception or production can be such an input. An introspective, linguistically aware person can readily compose a "synthetic" linguistic representation according to some arbitrary criterion: the first five words he can think of that begin with /b/, for example. This is obviously a very partial representation: just a sequence of phonological forms drawn from the lexicon, without explicit phonetics or syntax. But if this sequence is rehearsed, the phonetic level, together with whatever syntactic structure or traces of meaning may be accidentally implicit in the sequence, will be computed, just as if the sequence were what remained of a natural representation resulting from an earlier act of production, perception, or rehearsal. All that is required for a synthetic representation to serve as input for computing a natural one is that it contain enough information so that the rest of the structure of the utterance is more or less determined.

These various considerations suggest how it is that one linguistically aware language user can communicate with another, not by means of speech, but by means of synthetic representations, provided a way of transcribing such representations, that is, an orthography, is available. The writer speaks some utterance (at least to himself), creating a linguistic representation. The orthography enables him to transcribe this representation in some very partial fashion. From this transcription, the reader constructs a partial, synthetic linguistic representation. Such a representation is enough to enable the reader's linguistic processor to compute a complete, natural representation, as well as the writer's intended message.

If we compare what happens between writer and reader with what happens between speaker and hearer, it can be seen that the difference is much more than merely a matter of sensory modality. In speech perception, there is a natural and unique set of "signs"—the acoustic events that the human vocal tract can produce—and they are already in a form suitable for immediate linguistic processing (Liberman, this volume). Only the output of this processing is a linguistic representation. The input speech signal is in no sense a partial linguistic representation, but rather a complete representation of a very different kind. Moreover, the specification of the complex relation between the phonetically significant events in the signal and the units of the linguistic representation is acquired precognitively (Liberman & Mattingly, 1991); it does not have to be learned. Indeed, as has been remarked, the hearer has no access to the acoustic events, and may have little or no awareness of the units of the linguistic representation. In reading, on the other hand, there is no one, natural set of input symbols. Linguistic processing must therefore be preceded by a stage having no counterpart in speech perception: a cognitive translation from the orthographic signs to the units of the synthetic linguistic representation. The beginning reader must therefore deliberately master the mapping between the signs and the units, and for this he must have an awareness of the appropriate aspects of the linguistic representation.

CONSTRAINTS ON ORTHOGRAPHIC FORM

What psychological factors constrain the form of an orthography? Gelb (1963) makes a useful distinction between "outer form"—the shape of the visible symbols and their arrangement in a text—and "inner form"—the nature of the correspondence of the symbols to linguistic units. Beyond the trivial requirement that the symbols be visually discriminable, there appear to be no particular psychological constraints on outer form. The shapes of the signs in the writing systems of the world and the way they are arranged are extremely various, and such limitations as exist are to be accounted for not by cognitive or linguistic factors but by practical ones, such as the nature of the writing materials available and what patterns are easily written by hand, or by esthetic ones, such as the beauty of particular stroke patterns. This variety is possible because, as has just been seen, a cognitive translation is required for reading and writing in any event. This price having been paid, outer form can vary almost without limit.

Inner form, on the other hand, is highly constrained. In the first place, the orthography must correspond to the linguistic representation, because there is no other cognitive path to linguistic processes. This is the reason that proposals to treat spectrographic displays of speech as, in effect, an orthography the deaf could learn to read (Potter, Kopp, & Kopp, 1966) are not likely to succeed. On the one hand, the reader of spectrograms cannot process the visually-presented spectral information as a listener can process the same information in the auditorially-presented and biologically-privileged speech signal. On the other hand, the spectrogram reader has no natural cognitive access to raw spectral events, and, a fortiori, no awareness of them. Therefore, even if he could somehow synthesize a cognitive spectral representation from the visible one, there is no reason to believe it could be an input to linguistic processes. All he can do is to apply his cognitive knowledge of acoustic phonetics to the task of inferring the linguistic representation from the spectrogram. Because the relation between spectral patterns and even the most concrete level of this representation, the phonetic level, is extremely complex, and a great deal of extraneous information is present, "reading" spectrograms is a slow and unreliable process. Analogous observations, obviously, could be made with respect to other records of physical activity in which linguistic information is implicit, such as the speech waveform or traces of articulatory movements. What has to be transcribed, then, is some level or levels of the linguistic representation itself.

However, certain levels of the linguistic representation are seldom or never transcribed in traditional orthographies. For example, syntactic structure is never transcribed. The few features of orthography that might be considered syntactic, such as punctuation and sentence-initial capitalization, are more reasonably regarded as transcriptions of prosodic elements. Why is syntax thus avoided? It is not just that tree diagrams are cumbersome to draw and nested brackets difficult to keep track of, but that the syntactic structure alone would be insufficient to specify a particular sentence: Each possible phrase marker is shared by an indefinitely large number of sentences. It would therefore be necessary that a syntactic orthography also transcribe in some way the particular lexical choices. But if this is to be done, the phrase-marker itself becomes redundant,

because (barring some well-known types of structural ambiguity, such as those discussed by Chomsky, 1957) the words, and the order in which they occur, are themselves sufficient to specify syntactic structure.

Again, someone who supposed that speech and writing converged at the lowest conceivable level, given the difference of modality, might expect that the most efficient form of writing would be a narrow phonetic transcription (see Edfeldt, 1960). This transcription would correspond to the output of the phonological component of the grammar, presumably the level of the linguistic representation closest to the speech signal itself. Owing to contextual variation, higher-level units such as phonemes, syllables, morphemes, or words are not consistently transcribed or explicitly demarcated in such a transcription. But, in contrast to the syntactic orthography just considered, more than enough linguistic information to specify the linguistic representation would nevertheless be implicit. Why is such an orthography not found? A partial answer is that because, as has been suggested, writing and speech are not, in fact, so simply related, there is no particular advantage to a low-level, phonetically veridical representation. Moreover, it seems more difficult to attain awareness of phonetic details insofar as they are predictable. Once the language-learner is able to represent words phonemically, the phonetic level seems to sink below awareness. But as will be seen, there is a still more fundamental reason why a narrow phonetic transcription would be impractical.

It is important to distinguish between the linguistic unit used for the actual processing of an utterance by writer and reader, and the linguistic units to which the various graphemic units correspond. Elementary graphemic units correspond to phonemes (English letters or digraphs), syllables (Japanese kana[5]), or morphemes (simple Chinese characters). These are usually organized into complex units that have been called "frames" (Wang, 1981). A spelled word in English, a complex Chinese character, a grouping of Egyptian hieroglyphics are examples. Frames are usually demarcated by spaces in modern writing, but other demarcative symbols have been used. Sometimes the frame is implicit: The structure of the frame itself may be sufficient to demarcate it from adjacent frames, as in Japanese, where a kanji logogram or logograms is regularly followed by kana syllable signs specifying affixes. Some orthographies, such as those early alphabetic orthographies in which there is no demarcative information of any kind, have no frames larger than their elementary signs. Frames often correspond to linguistic words, but not always: In Chinese and Sumerian, they correspond to morphemes.

By "unit of transcription" is meant the linguistic unit that the writer actually transcribes and the reader cognitively translates to form the synthetic linguistic representation. One might expect that the units of transcription for a particular orthography would be those to which its frames corresponded. Thus, in English, the frames are consistent spellings of words, and the experienced reader's intuition is surely that he reads word by word and not letter by letter, as he would if the transcription unit were the segment. This intuition is borne out by demonstrations of "word superiority." In these experiments, it is found, for example, that subjects can recognize a letter faster and more accurately when it is part of a real written word than when it appears alone or in a nonword (Reicher, 1969). This result suggests that in the case of a real word, subjects can use the orthographic information to recognize the word very rapidly, and then report the letters it contains. If the segment were the transcription unit, the letters corresponding to the segments should be recognized and reported faster than the words.

However, it is possible that the unit of transcription does not really depend on the frame used in a particular orthography, but is in fact *always* the word. One reason for believing this is that the word has to be the most efficient unit of transcription, because words are the largest lexical structures. Anything smaller would require processing more units per utterance; anything larger could not be readily coded orthographically.

Chinese writing allows a test of this possibility. A Chinese word consists of one or more monosyllabic morphemes. In the writing, characters are the frames and correspond to these morphemes. Words as such are not demarcated. There is some evidence, however, that the unit of transcription is nonetheless the word. In a recent experiment (Mattingly & Xu, in preparation), Chinese speakers were shown sequences of two characters on a CRT. In half the sequences, one of the characters was actually a pseudocharacter, consisting of two graphic components that in actual writing occur separately as components of other characters, but not together in the same character. Of the sequences in which both characters were real, half were real bimorphemic words and half were pseudowords. The subject's

task was to respond "Yes," if both characters in a sequence were genuine and "No," if either was a pseudocharacter. Subjects performed this task faster for words than for pseudowords, and it was possible to show that this was not simply an effect of the higher transitional probabilities of the word sequences, but rather a valid "word superiority" effect. This result, like that of an earlier experiment by C. M. Cheng (1981, summarized in Hoosain, 1991) suggests that despite morphemic framing and the absence of word boundaries, the word is the transcription unit for Chinese readers. Other writing systems in which words are not framed remain to be investigated.

But if word-size frames are not essential for reading word by word, why is a narrow phonetic transcription an unlikely orthography? The reason must be that the shapes of words in such a transcription are context-sensitive and thus difficult to recognize. (Notice what happens to /hænd/, *hand*, in [hæntuwlz], *hand tools*, [hæŋgrənejd], *hand grenade*, [hæmpɪkt], *hand picked*, etc.). The reader is therefore forced to process the transcription symbol by symbol, a slow and arduous procedure. In Chinese, on the other hand, though word-boundaries are absent, the form of an orthographic word is constant, or at least not subject to contextual variation. It is suggested that this is a minimal constraint that all writing systems must meet, so that words can serve as units of transcription.

Although words are the transcription units, writing always employs graphemic units corresponding to linguistic units smaller than the word. It might seem possible, in principle, to have a pure logographic system, consisting simply of one monolithic symbol for each word. But the difficulty with such a system is that while the lexicon of a language is, in principle, finite, it is in practice, indefinite: New words are continually being coined or borrowed. In some cases—a nonce word or an unusual foreign name, for example—it would make little sense to provide a special logogram. A writer could thus find himself with no means of writing a particular word because no logogram for it existed. Or, of course, he could be stuck simply because he did not know the correct logogram. An actual writing system insures that the writer will never be in this situation by providing a system of spelling unit The availability of the spelling system guarantees that the orthography will be "productive," that is, that the writer who has mastered the spelling rules will always have some way (though it may not be the "correct" or standard way) to write every word in the language (Mattingly, 1985).

The only linguistic units that have served as the basis for spelling units are syllables and phonemes. It might be thought that morphemes could be the basis of a spelling system and some (e.g., Sampson, 1985) have argued that Chinese has such a system, because the characters correspond to morphemes. This is true, but, as has already been noted, these morphemic units are frames: Relatively few of the characters in the inventory are simple logograms. Over 90% are phonetic compounds, each consisting of two graphic components that (in general) occur also as separate logographic characters. One of these, the "phonetic" stands, in principle, for a particular phonological syllable, and the set of phonetics thus constitutes a syllabary. The other, the "semantic," is one of 214 determiners that serve to mitigate the extensive homophony of Chinese: The number of monosyllabic morphemes far exceeds the number of phonologically distinct syllables. The situation is complicated, however, because there is usually more than one phonetic corresponding to a particular phonological syllable (there are about 4000 in all for about 1300 phonologically distinct syllables), and because, through various accidents of linguistic history, a phonetic often has different phonological values in different characters. But these circumstances should not obscure the highly systematic, syllabographic nature of the spelling, any more than the existence of several spelling patterns for one sound, and numerous inconsistencies in letter-to-sound correspondence, should obscure the systematic, alphabetic nature of English spelling (DeFrancis, 1989).

Words can indeed be analyzed into morphemes as well as segments and syllables, but the inventory of morphemes in a language, like the inventory of words itself, is indefinitely large and subject to continual change. While logograms that are morphemic signs can have a valuable supplementary function in orthography, they could not constitute a productive spelling system, and there is no orthography in which they play this role.

Syllables and segments, on the other hand, have several properties that make them suitable as a basis for spelling units. First, a word can always be analyzed as a sequence of phonological elements of either type. Second, the inventory of syllables may be small (and indeed was small in all the languages for which syllabic spelling

developed independently) and the inventory of segments is *always* small. Third, the membership of these inventories changes only very slowly. No other linguistic units have these convenient properties, save perhaps phonological distinctive features (Because a diacritic is used to indicate voicing, it could be maintained that features have a marginal role in Japanese spelling).

In sum, every orthography needs to have a spelling system and a spelling system is necessarily phonographic. It is not accidental that all orthographies spell either syllabically or segmentally: there is probably no other way to spell.

THE INVENTION OF WRITING[6]

Writing was invented, probably several times, by illiterates. From what has been said already, it follows that what had to be discovered was one or the other of the two possible spelling principles, the syllabic or the segmental, and that this must have required awareness of these units of the linguistic representation. How could the inventors have arrived at such awareness?

Some linguistic units seem to be more obvious than others. Awareness of words can perhaps be assumed for most speakers, even if they are preliterate or illiterate. It probably requires only a very modest degree of awareness to appreciate that an utterance is analyzable as a sequence of syntactically functional phonological strings, if only because sequences consisting of just one such string are quite frequent: Words may occur in isolation. Certainly preliterate children have no difficulty in understanding a task in which they are to complete a sentence with some word, and a linguist's naive informant readily supplies the names of objects. Awareness of syllables as countable units may also be fairly widespread. The syllable is the basis for verse in many cultures; preliterate children can count the number of syllables in a word. This kind of syllabic awareness, however, is probably not the same thing as being aware (if such is indeed the case) that the syllables of one's language constitute a small inventory of readily demarcatable units.

These limited degrees of linguistic awareness are probably readily available to speakers of all languages. But more subtle forms of awareness may well have arisen only because they were facilitated by specific properties of certain languages, including, in particular, those for which writing was originally invented.

Consider, first, Chinese. In the Ancient Chinese language, words were in general monomorphemic, there being neither compounding nor affixation. Morphemes were monosyllabic and a particular morpheme was invariant in phonological form. Because of restrictions on syllable structure, the inventory of syllables was small. Homophony was therefore very extensive, one syllable corresponding to many morphemes (Chao, 1968).[7] The number of different characters in the Chinese writing system sharing a particular phonetic component gives some notion of the degree of homophony in Ancient Chinese, and this number often exceeds twenty. Chinese thus contrasts sharply with English and other Indo-European languages, in which morphemes vary in phonological form, may be polysyllabic, and may not even consist of an integral number of syllables; syllable structure is complex; the number of possible syllables is relatively large; and homophony is therefore a marginal phenomenon.

Since words coincided with morphemes in Chinese, awareness of morphemes required no analysis, and the use of logograms, i.e., morphemic signs, was an obvious move. The extensive homophony made "phonetic borrowing"—using the sign for one morpheme to write another morpheme with the same syllabic form[8]—a strategy that was both obvious and productive; when a writer needed to write a morpheme, a sign with the required sound was very likely to be available. It thus became obvious that the number of different sounds was in fact small, yet every morpheme corresponded to one of them. Awareness of demarcatable syllable units thus developed. Of course, the same extensive homophony that fostered the discovery of these units also meant that their signs had to be disambiguated by the use of logograms as determiners, as in the large class of characters called "phonetic compounds," described earlier.

Chinese morphophonological structure thus encouraged the discovery of the syllable; on the other hand, it did not encourage the discovery of the phonemic segment. There was nothing about this structure that would have served to isolate phonemes from syllables or morphemes.

Sumerian was an agglutinative language. A word consisted of one or two monosyllabic CVC morphemes and various inflectional and derivational affixes. Its phonology had certain properties that imply a preference for a CVCVC...VC syllabification. There were no intrasyllabic consonant clusters; a cluster simplification process deleted the first of two successive consonants across syllable boundaries, resulting in such alternations as til, ti, 'life'; and final vowels were deleted (Driver,

1976; Kramer, 1963). In other relevant respects, however, Sumerian resembled Chinese and, like Chinese, favored awareness of morphemes and of syllables as demarcatable units. Aside from the effects of the syllable-forming processes just mentioned, a root maintained an invariant phonological form. A root could be repeated to indicate plurality. Because the morphemes were monosyllabic, and because of the restricted syllable structure, the number of possible distinct syllables was small. These circumstances, resulted, again, in extensive homophony.

For a speaker of Sumerian to become aware of morphemes was perhaps not quite as easy as for a speaker of Chinese. He would have had to notice that words with similar meanings often had common components, for the most part corresponding to syllables. This stage of awareness having been achieved, morphemic writing is possible. From this point on, the story is quite similar to that for Chinese, homophony leading to phonetic borrowing, and then to syllable writing supplemented with determiners.

There is, however, one striking difference between the Sumerian and the Chinese writing systems. While Chinese makes no internal analysis of syllables, Sumerian does. A sign for a $C_1V_1C_2$ morpheme could be borrowed to write a $C_1V_1C_3$ morpheme, e.g., the RIM sign was used to write **rin**. A VC syllable sign could be used as a partial phonetic indicator after a logogram, e.g., GUL + UL. For many of the $C_1V_1C_2$ syllables, as has been mentioned, there was no special sign; instead, such a syllable was written with the sign for the C_1V_1 followed by the sign for V_1C_2. Thus the syllable **ral** is written RA AL (examples from Gelb, 1963). A possible explanation of these various practices is that in spoken Sumerian, consistent with its preference for CVCVC...VC structure, some form of vowel coalescence took place when two similar vowels came together, so that $C_1V_1 + V_1C_2$ sequences became phonetically $C_1V_1C_2$, and thus homophonous with original $C_1V_1C_2$ syllables. Such homophony could have suggested analyzing and so writing the latter as $C_1V_1 + V_1C_2$. Again CV signs as well as VC signs were used to indicate the endings of $C_1V_1C_2$ morphemes. For example, because of multiple semantic borrowing, the logogram DU could stand not only for **du**, 'leg,' but also for **gin**, 'go,' **gub**, 'stand,' and **tum**, 'bring'. Which of the latter three was intended was indicated by writing DU NA for **gin**, DU BA for **gub**, and DU MA for **tum** (Driver, 1976). This practice perhaps arose because the phonological final vowel deletion made $C_1V_1C_2$ and $C_1V_1C_2V_2$ sequences homophonous, suggesting that what followed C_1V_1 could be written in either case as if it were C_1V_2. Thus the Sumerians may have viewed $C_1V_1C_2$ morphemes either as $C_1V_1 + V_1C_2$ or as $C_1V_1 + C_2V_2$, either of which was entirely consistent with their syllabic phonological awareness.

With Egyptian, in contrast to Chinese and Sumerian, the morphology and phonology of the language of the language favored segmental awareness. In Afro-Asiatic languages, the roots are biconsonantal and triconsonantal patterns into which different vowels or zero (that is no vowel at all) are inserted to generate a large number of inflected forms. Because the vowels of Egyptian are unknown, it is easier to illustrate this point with an example from another Afro-Asiatic language, e.g., Hebrew. From the Hebrew root k-t-b are derived **kātab**, 'he wrote'; **yikkātēb**, 'he will be inscribed'; **kātob** 'to write'; **kātub**, 'written'; **miktab**, 'letter; and many other forms. Because of phonological restrictions, the number of different consonantal patterns in Egyptian was relatively small, and there were consequently numerous homophonous roots, e.g., **n-f-r**, 'good'; **n-f-r**, 'lute' (Jensen, 1970).

It is not difficult to imagine an Egyptian noticing that many sets of semantically similar words in his language had a common consonantal ground and a varying vocalic figure, though at first he may not have individuated the consonants. Accordingly, signs for root morphemes were devised. The homophony of Egyptian then did for phonetic segments what homophony in Chinese and Sumerian did for syllables. A morphemic sign was frequently borrowed to write a homophonous morpheme, e.g., NFR, the sign for **n-f-r**, 'lute', used to write **n-f-r**, 'good,' or WR, 'swallow,' used to write **w-r**, 'big.' The signs were now generalized to stand for consonantal sequences that were not morphemes, e.g., WR < WR was used to write the first part of **w-r-d**, 'weary.' And because in some cases roots were actually uniconsonantal, and in other cases the second consonant had become silent, some signs came to stand for single consonants, and constituted a consonantal alphabet. Thus the **d** in **w-r-d** could written with the sign D < DT, the final consonant in **d-t**, 'hand,' being actually the feminine suffix, not part of the root. Finally, logograms were employed as determiners to clarify ambiguous transcriptions: the spelling MN N H for the word **m-n-h** being followed by the determiner for 'plants' when this word had the sense 'papyrus plant,' the determiner for 'men'

when it had the sense 'youth,' and the determiner for 'minerals' when it had the sense 'wax' (examples from Jensen, 1970). In this fashion, the Egyptians arrived at a consonantal spelling system.

If the Egyptians had thus achieved segmental awareness, why did they not transcribe the vowels as well as the consonants? It is not likely that they were unable to hear the different vowels. The explanation is rather that because the vowels ordinarily conveyed only inflectional information, the writing was sufficiently unambiguous without such indications, just as English writing is sufficiently unambiguous without stress marking. But as has already been noted, there was a convention for writing vowels when necessary. Such writing is found very early in the history of Egyptian writing (Gelb, 1963).

The Egyptians could hardly have arrived at a syllabic system instead. Because zero alternated with vowels in the generation of words, there was no obvious correspondence between morphemes and syllables or syllable sequences. And because of such alternations, a syllabic orthography would have resulted in a number of dissimilar spellings for the same morpheme.

These examples suggest that the phonological awareness required for the invention of writing develops when morphemes have a highly restricted phonological structure—monosyllabic, in the case of Sumerian and Chinese; consonantal in the case of Egyptian—that results in pervasive homophony. Speakers of such languages are naturally guided to the invention of writing by these special conditions. (A corollary is that it is not necessary to propose a derivation of Egyptian from Sumerian to account for parallels in the development of the two systems.) On the other hand, Indo-European languages and many others lack any such restrictions, and would not have favored phonological awareness in this way. Indeed, one has to wonder whether, for such languages, writing could have been invented at all.

In the early discussion of the psychology of reading, the precise role of phonological awareness in learning to read appeared equivocal. Is phonological awareness a prerequisite for reading? Or, on the other hand, does the experience of reading engender phonological awareness (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977)? It was later seen, however, that both statements must be true: The beginning reader must, indeed, have some degree of awareness, but this awareness is increased and diversified in appropriate directions as a result of his encounter with the orthography (Morais, Alegria & Content, 1987). In the same way, the invention of writing must have been an incremental process, beginning with an initial awareness of morphemic structure. The experience of working out ways to transcribe morphemes for which there were no logograms led to awareness of the syllabic or phonemic structure of these morphemes, and then to awareness of such structure generally.

To say that the process was incremental is not to say that it was not quite rapid. It is noteworthy that in all three of the writing traditions just considered, evidence of spelling is found very early: in Sumerian writing from the Uruk IV stratum (Gelb, 1963); in Chinese writing of the Shang dynasty (DeFrancis, 1989); in Egyptian writing of the First Dynasty (Gelb, 1963). These facts are consistent with the proposal that for general-purpose writing, a purely logographic system is impractical. As has been argued, an orthography is not productive without a spelling system: The invention of the one requires the invention of the other.

To the extent that this account of the invention of writing is plausible, it supports the dichotomy between syllabic and segmental spelling proposed earlier, for what had to be invented was one or the other of the two spelling principles that provide the basis for the classification. It should also be noted that the segmental principle did not develop in Egypt by elaborating on the syllabic principle, but rather by generalizing from the segmental transcription of morphemes: The syllable played no role. And, conversely, when Sumerians analyzed complex syllables, they did not resolve them into their constituent phonemes, but rather into simpler syllables. The discovery of one method almost seems to have guaranteed that the other would not be discovered. In effect, speakers of these languages come to regard them as as essentially syllabic or as essentially segmental, and their writing systems reflect one of these two phonological theories.

TRANSMISSION OF WRITING SYSTEMS

It has already been noted that orthographic traditions are either consistently syllabic or consistently segmental. Some explanation for this consistency is required. It seems natural enough, perhaps, that a segmental tradition should not become syllabic, for this would appear to be a backward step. But that no syllabic tradition

should have become segmental is puzzling, the more so because there have been at least two occasions when such a development might reasonably have been expected. The first was when speakers of Akkadian, an Afro-Asiatic language with consonantal root structure similar to that of Egyptian and Hebrew, borrowed Sumerian syllabic writing. A proper awareness of the morphophonology of their language would have suggested that they convert the Sumerian system into a consonantal system. But instead, the Akkadians preserved the syllabic character of the borrowed writing, even though to write the same triconsonantal pattern in different ways depending on the particular inflectional vowels obscured the roots of native words. Similarly, the Mycenaean Greeks borrowed Minoan syllable writing, and instead of making an alphabet out of it, as would have been sensible, given the extensive consonant clustering in Greek, they continued to write with signs that stood for CV syllables, either ignoring the "extra" consonants or pretending that they were syllables. This resulted in such bizarre transcriptions such as A RE KU TU RU WO for **alektruōn** 'cock' (Ventris & Chadwick, 1973). What can have happened to linguistic awareness in these cases?

The explanation begins with the observation that the mismatches between language and writing observed for Akkadian and Mycenean Greek are not unparalleled; they are simply fairly extreme cases. While an originally invented writing system clearly reflects the morphophonological structure of the language it was invented to write, this situation is obviously exceptional. In general, the system used at a particular time to write a particular language has been inherited from an earlier stage in the history of that language, or has been adapted from a system (itself perhaps an adaptation) used for some other language, or, most commonly, both. The consequence, in many cases, is that the writing often seems very poorly suited to the spoken language. If Akkadian and Mycenaean Greek illustrate the risks of borrowing, the English writing system is a good illustration of the effects of orthographic inheritance. The phonology of English has changed considerably since the fifteenth century, most notably in consequence of the Great Vowel Shift, but the writing system has remained very much as it was then (Pyles, 1971). As a consequence, the system has a number of features that must seem very peculiar to the foreigner learning English: For example, the same letter is used to write phonetically dissimilar vowels, a tense vowel is denoted by an E after the following consonant, and a lax vowel is denoted by the doubling of this consonant. A similar account could be given for Chinese writing, which corresponds more closely to Classical Chinese than to any modern dialect.

It cannot be doubted, given what has been learned in recent years about the relation between orthographic structure and learning to read in modern languages, that such complications place a heavy burden on the learner (Liberman, Liberman, Mattingly, and Shankweiler (1980). What is surprising, given the close connection between literacy and awareness of linguistic representations, a connection clearly essential in the invention of writing, is that readers and writers have so often happily accepted (once they have learned it) an orthography that seems poorly matched to their language. It might have been expected that Akkadian cuneiform would have been rejected as soon as it was proposed, and that English orthography would by now have been abandoned as obsolete. But, instead, it is reported that the Akkadians believed their writing system to be of divine origin (Driver, 1976), and Chomsky and Halle (1968) say that "conventional [English] orthography is...a near optimal system for the lexical representation of English words" (p. 49).

In the case of inherited orthographies, the explanation may be that the orthography itself may determine not only which aspects of linguistic representations are singled out for awareness, but perhaps, indirectly, the character of these representations themselves. This could come about if the orthographically based, synthetic input representations were taken seriously by the language processor as evidence about the structure of the language, and thus led to adjustments in the beginning reader's morphophonology. It will be recalled that according to the sketch of the reading and writing process given earlier, the processor does not distinguish synthetic representations from natural ones. Consistent with this possibility is the fact that orthographic conventions sometimes mimic phonology: The conventions for marking English tense and lax vowels invite the reader to assume that underlying lax vowels become tense in open syllables and underlying tense vowels become lax before underlying geminate consonants. Such pseudophonological rules, as well as derivational morphological relations as those between *heal, health* or *telegraph, telegraphy,* though at first having merely orthographic status, may acquire linguistic reality for the experienced reader.[9] For such a reader, the orthography corresponds to linguistic

representations because the representations themselves have been appropriately modified, and English orthography now indeed seems "near optimal."

In the case of borrowed orthographies, a similar explanation may apply. The phonological awareness of a borrowing group, such as the Akkadians or the Greeks, was not guided by peculiarities of their own spoken language, as was the awareness of the original inventors of writing, but by the writing system they were borrowing. This is hardly surprising: The borrowers were not sophisticated consumers, comparing competing technologies to decide which was better for their particular needs. They did not realize that there was a choice that could be made between the two different spelling principles and the theories of phonology implicit in each. They simply embraced unquestioningly the spelling principle—syllabic in the cases considered above—used by the culture under whose influence they had come, just as beginning readers accept the principle of the writing system they inherit. This principle having been accepted, the morphophonologies of the borrowers adjusted so that their linguistic representations became, in fact, a good match to their syllabic orthographies.

If this account is correct, it has to apply to the transmission of segmental systems, as well. A segmental system has obvious advantages over a syllabary for languages with complex syllable structure. But the spread of the alphabet is perhaps to be explained by an appeal to the forces of tradition rather than to those of reason.

An orthographic tradition can perpetuate itself because it offers a particular brand of morphophonological awareness ready-made. The processes of introspection needed to invent writing in the first place are not demanded. The kind of awareness offered may be poorly matched to a particular language, but this does not impede the process. Whether the writing system is borrowed or inherited, the morphophonology of the new reader adjusts to meet the presuppositions of the system.

CONCLUSIONS

It has for some time been widely agreed that the notion of linguistic awareness is essential for an understanding of the reading process, the acquisition of reading and reading disability. This notion is likewise essential for an understanding of the invention and dissemination of orthographies. There are really only two possible ways to write, the syllabic method and the segmental method, because only by using one of these two methods is the writer assured of being able to write any word in his language. But for an illiterate to discover either of these methods, and thus be in a position to invent writing, requires awareness of the appropriate unit of linguistic representations. Awareness of syllables, or, on the other hand, of segments, is fostered by special morphophonological properties found in those languages for which writing systems were invented, though by no means in all languages. But once it has become established, the writing system itself shapes the linguistic awareness, and even the phonology, both of those who inherit the system and of those who borrow it to transcribe some other language. Thus, in the history of writing, syllabic and segmental traditions are clearly distinguished.

REFERENCES

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Chao, Y.-R. (1968). *Language and symbolic systems*. London: Cambridge University Press.

Cheng, C. M. (1981). Perception of Chinese characters. *Acta Psychologica Taiwanica*, 23, 137-153.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cross, F. M., & Freedman, D. N. (1952). *Early Hebrew orthography: A study of the epigraphic evidence*. New Haven: American Oriental Society.

DeFrancis, J. (1950). *Nationalism and language reform in China*. Princeton, NJ: Princeton University Press.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Driver, G. R. (1976). *Semitic writing: From pictograph to alphabet*. London: Oxford University Press.

Edfeldt, A. W. (1960). *Silent speech and silent reading*. Chicago: University of Chicago Press.

Edgerton, W. F. (1952). On the theory of writing. *Journal of Near Eastern Studies*, 11, 287-290.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Gelb, I. J. (1952). *A study of writing: The foundations of grammatology*. Chicago: University of Chicago Press.

Gelb, I. J. (1963). *A study of writing* (rev. ed.). Chicago: University of Chicago Press.

Hill, A. A. (1967). The typology of writing systems. In W. M. Austin (Ed.), *Papers in linguistics in honor of Leon Dostert* (pp. 92-99). The Hague: Mouton.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum.

Hyman, L. (1975). *Phonology theory and analysis*. New York: Holt-Rinehart-Winston.

Jensen, H. (1970). *Sign, symbol and script: An account of man's efforts to write* (3rd ed.). Tr. G. Unwin. London: G. Allen & Unwin.

Kelley, D. H. (1976). *Deciphering the Mayan script*. Austin, TX: University of Texas Press.

Kramer, S. N. (19[illegible]). *The Sumerians: Their history, culture, and character*. Chicago: University of Chicago Press.

Liberman, A. (1992). The relation of speech to reading and writing.

Liberman, A. M., & Mattingly, I. G. (1991). Modularity and the effects of experience. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: applied and ecological perspectives* (pp. 33-38). Hillsdale, NJ: Lawrence Erlbaum.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. Venezky (Eds.), *Orthography, reading, and dyslexia* (pp. 137-153). Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, W. F. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. A. Scarborough (Eds.), *Towards a psychology of reading* (pp. 207-225). Hillsdale, NJ: Lawrence Erlbaum.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly, *Language by ear and by eye: The relationships between speech and reading* (pp. 133-147). Cambridge, MA: MIT Press.

Mattingly, I. G. (1985). Did orthographies evolve? *RASE Remedial and Special Education, 6(6)*, 18-23.

Mattingly, I. G. (1987). Morphological structure and segmental awareness. *CPC cahiers de psychologie cognitive, 7*, 488-493.

Mattingly, I. G. (1991). Reading and the biological function of linguistic representations. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 339-346). Hillsdale, NJ: Lawrence Erlbaum.

Mattingly, I. G., & Xu, Yi (in preparation). *Word superiority in Chinese.*

Morais, J., Alegria, J., & Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: An interactive view. *CPC Cahiers de Psychologie Cognitive, 7*, 415-438.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Potter, R. K., Kopp, G. A., & Kopp, H. G. (1966). *Visible speech* (2nd ed.). New York: Dover.

Pyles, T. (1971). *The origins and development of the English language* (2nd ed.). New York: Harcourt Brace Jovanovich.

Ray, J. D. (1986). The emergence of writing in Egypt. *World Archaeology, 17(3)*, 308-316.

Read, C., Zhang, Y.-F., Nie, H.-Y., & Ding, B.-Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition, 24*, 31-44.

Reicher, G. M. (1969). Perceptual recognition as function of the meaningfulness of the stimulus material. *Journal of Experimental Psychology, 81*, 275-280.

Sampson, G. (1985). *Writing systems A linguistic introduction.* Stanford, CA: Stanford University Press.

Ventris, M., & Chadwick, J. (1973). *Documents in Mycenaean Greek* (2nd ed.) Cambridge: Cambridge University Press.

Wang, W. S.-Y. (1981). Language structure and optimal orthography. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp. 223-236). Hillsdale, NJ: Lawrence Erlbaum.

FOOTNOTES

*In L. Katz & R. Frost (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 1-16). Amsterdam: Elsevier Science Publishers (1992).

†Also University of Connecticut, Storrs.

1 It will be assumed here, following Gelb (1963), Jensen (1970), DeFrancis (1989) and others, that there are six major orthographic traditions: (1) Mesopotamian cuneiform, beginning with Sumerian (c. 3100 B.C.) and including Akkadian, cuneiform Hittite, Urartian, Hurrian, Elamite, Old Persian; (2) Cretan, including Minoan Linear A, Mycenaean Greek Linear B, Cypriote, and Hittite hieroglyphics, all probably derived from a common source (c. 2000 B.C.); (3) Chinese, beginning with Chinese itself (c. 1300 B.C.) and including Korean nonalphabetic writing and Japanese; (4) Mayan (c. 300 A.D.); (5) Egyptian (c. 3000 B.C); (6) West Semitic, beginning with Phoenician (c. 1600 B.C.) and including Ras Shamrah cuneiform, Old Hebrew, South Arabic, Aramaic, and Greek alphabetic writing. From Aramaic derive Hebrew, Arabic, and many others; from Greek derive Etruscan, Latin, and many others. Germanic runes and Korean alphabetic writing probably belong in this tradition also, though the derivations are not clear. All but the most dogmatic monogeneticists would agree that the Mesopotamian, Cretan, Chinese, and Minoan traditions are probably independent developments. But some scholars (e.g., Driver, 1976; Ray, 1986) would derive Egyptian writing from Mesopotamian, and some (e.g., Driver, 1976), with somewhat greater plausibility, would derive West Semitic from Egyptian.

2 Egyptologists and most other students of writing believe that Egyptian phonographic signs stand for consonants, the vowels not being regularly transcribed. But according to Gelb, they stand instead for generalized syllables, e.g., the Egyptian sign usually interpreted as consonantal **w** actually stands for **wa, wi, we, wu,** or **wo,** according to context. It is obviously difficult to distinguish these two accounts empirically. The only support Gelb offers for his position is that "the development from a logographic to a consonantal writing, as generally accepted by the Egyptologists, is unknown and unthinkable in the history of writing" (Gelb 1963, p. 78). But this argument is clearly circular (Edgerton, 1952; Mattingly,1985).

3 Gelb (1952, 1963) proposed some cases in which syllabic systems are supposed to have developed into segmental systems ; but see Edgerton (1952). Ethiopic writing, derived from the West Semitic consonantal tradition, might be viewed as a syllabic system derived from a segmental system, because the signs do correspond to syllables. But, with a few exceptions, each sign actually consists of a consonant letter plus a vowel mark, except that **a** is left unmarked. As in the case of Indic systems, one could argue about whether this is a consonantal or a *plene* system, but it is certainly not a syllabic system (Sampson, 1985).

4 The proposals in this section are developed in more detail in Mattingly (1991).

5 Japanese kana correspond, strictly speaking, to moras, which are not equivalent to English syllables. But they do belong to a general class of phonological units that can be called "syllables" (see, e. g., Hyman, 1975).

6 An earlier formulation of some of the proposals in this section can be found in Mattingly (1987).

7 DeFrancis (1950), protesting against the "monosyllabic myth," has suggested that there actually were many polysyllabic words in Ancient Chinese, just as in Modern Chinese, but that only one of the syllables in a word was transcribed in the writing. Thus, morphemes that appear from the writing to be monosyllabic homophones may actually have been polysyllabic morphemes with common homophonous syllables. Y.-R. Chao's (1968) response was that "so far as Classical Chinese and its writing system is concerned, the monosyllabic myth is one of the truest myths in Chinese mythology" (p. 103). For the present purpose, however, it does not matter whether the myth is true or false. DeFrancis's partial homophony will serve as well as the total homophony more usually attributed to Ancient Chinese.

8 Or, on DeFrancis' (1950) view, another morpheme having a syllable in common.

9 These changes in the morphophonologies of individual readers have, by hypothesis, no basis in the spoken language and are transmitted only from writer to reader, and not from mother to child. Thus, though psychologically real, they are not part of the grammar of the language as usually conceived of.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 141-146

The Effects of Aging and First Grade School on the Development of Phonological Awareness*

Shlomo Bentin,† Ronen Hammer,†† and Sorel Cahan††

The independent influence of aging and schooling on the development of phonological awareness was assessed using a between-grades quasi-experimental design. Both schooling (first grade) and aging (5-7 years) significantly improved children's performance on tests of phonemic segmentation, but the schooling effect was four times bigger than the aging effect. The schooling effect was attributed to formal reading instruction, whereas the aging effect probably reflects natural maturation and informal exposure to written language. These data support a strong mutual relation between reading acquisition and phonological awareness.

Phonological awareness is the aptitude of being aware of the phonemic structure of spoken words. It is usually assessed by testing the subjects' ability to isolate and manipulate individual phonemic segments in words.

Although as soon as a child is able to understand and produce speech he obviously makes phonemic distinctions, the ability to manipulate phonemic segments consciously develops only around the first grade in the elementary school. For example, Liberman, Shankweiler, Fisher, and Carter (1974) found that none of the prekindergartners and only 17% of the kindergartners tested were able to parse words into phonemes, while 70% of the first graders tested succeeded in doing so.

The significant improvement in phonological awareness at this age may be primarily ascribed to one of two factors (which are not mutually exclusive): (1) cognitive-linguistic skills which mature at about the age of six independent of formal reading instruction (Bradley & Bryant, 1983); or (2) learning to read in an alphabetic orthography (Bertelson, Morais, Alegria, & Content, 1985). In contrast to speech, where individual phonemes are coarticulated and overlap in the acoustic stream, in writing the phonemes are represented by clearly defined orthographic segments, the letters (see Liberman & Mattingly, 1989). Assuming that children learn about these letter-sound correspondence when they learn to read, it seems likely that during the acquisition of reading skills they become explicitly aware that words are formed of the sounds which the letters represent. Owing to the impossibility to experiment with elementary school attendance, the effect of reading instruction on phonological awareness has been investigated only indirectly in studies that have relied on natural variation: (1) between literate and illiterate adults; (2) between different orthographic systems (alphabetic vs. logographic) among literates; or (3) in the emphasis upon letter-sound correspondence between reading instruction methods within the alphabetic system (e.g., "analytic" vs. "global" methods).

Most of these studies suggested that learning to read triggers, or at least promotes the development of phonological awareness. For example, Morais, Cary, Alegria, and Bertelson (1979) reported that the performance of illiterate adults on tests of phonemic segmentation was inferior to that of other adults from the same rural community who learned to read in adulthood (see

This study was supported by the Israel Foundations Trustees, and by the NCJW Research Institute for Innovation in Education, School of Education, the Hebrew University. We thank Lynette Bradley, Paul Bertelson, Paul Bryant, Al and Isabelle Liberman, Cyma Van Peten, and an anonymous reviewer for constructive comments. The cooperation of the principals, the teachers and the students of the schools and kindergartens of Jerusalem are highly appreciated.

also Morais, Castro, Scliar-Cabral, Kolinsky, & Content, 1987; Morais, Bertelson, Cary, & Alegria, 1986). In Chinese adults, Read, Zhang, Nie, and Ding (1986) found higher phonological awareness in subjects who learned to read the alphabetic (pinyin) orthographic system than in subjects who read only the logographic system (kanji). Equivalent results were found with children in first class; those who learned to read according to the "analytic" (segmental) method performed better on tests of phonemic segmentation than those who learned to read by the "global" (holistic) method (Alegria, Pignot, & Morais, 1982).

However, while the studies cited above suggest that literacy influences the development of phonological awareness they do not prove this claim. The caveat is that they all share the serious problem of possible confounding of differences in the extent or method of reading acquisition with other variables that may have influenced phonological awareness (e.g., the amounts of informal linguistic experience). Therefore, there is still a need to specify the effect of schooling in general and reading acquisition in particular, on to the sharp improvement in phonemic segmentation ability which occurs in the first year of schooling. Such a specification is important particularly because claims about the causal link between phonological awareness and literacy have been largely based on positive correlations found between the performance of children in tests of phonemic segmentation and their reading skills in English (e.g., Bradley & Bryant, 1985; Liberman, 1973; Fox & Routh, 1975; Treiman & Baron, 1981) as well as in other languages such as Italian (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988), Swedish (Lundberg, Olofsson, & Wall, 1980), Spanish (de Manrique & Gramigna, 1984), and French (Bertelson, 1987).

The present study circumvents the confounding problem by utilizing a recently introduced quasi-experimental paradigm, that allows for the post hoc disentangling of the independent effects of age and schooling (Cahan & Davis, 1987). This approach entails administration of the same test to at least two adjacent grade levels and takes advantage of the school cutoff that is imposed in most countries. The overall cross-sectional increase in mean test scores as a function of age is decomposed into within-grade and between-grades segments which can be attributed to age and schooling effects, respectively.

Theoretically, this could be achieved by comparing children born one day before the cutoff date with children born one day after (Morrison, 1988); those children will differ by only one day in age, but by a full year of schooling. Similarly, children that are born in the first and the last day of one schooling year will differ in age by a full year while being in the same grade. Unfortunately, aside of the logistic difficulty to find enough children in each birth date group, this approach suffers from a serious shortcoming of selection, because the cutoff date is never strictly imposed. Moreover, those exceptions are not random: Intellectually advanced children who are slightly younger than the official school age are often admitted, while children who are somewhat older than the cutoff point but insufficiently developed may be held back an additional year (Cahan & Davis, 1987, Cahan & Cohen, 1989). This creates a situation of "missing" children in each grade, particularly among children at the extreme age points. Such selective misplacement usually leads to overestimation of the schooling effect (Cahan & Cohen, 1989).

A possible solution of the selection problem is to base the estimation of age and schooling effects on the predicted (rather than empirically obtained) mean test scores of the youngest and the oldest children in each grade. Prediction would be based on the best fitting regression of test scores on chronological age across the entire legal age range in that grade, with the exclusion of the selection-tainted birth dates near the cutoff point. This idea underlies the recently proposed between-grades regression discontinuity design (Cahan & Davis, 1987). In the present study we applied the same model to the estimation of the independent effects of one year of schooling (during which reading acquisition was the primary curricular activity) and one year of aging on the development of phonological awareness as evidenced by tests of phonemic segmentation.

Method

Design. The "between-grades" quasi-experimental paradigm (Cahan & Davis, 1987) relies on two assumptions: (1) the "allocation" of children to birth dates is random, and (2) the grade level is solely a function of chronological age, that is admission to school is based only on chronological age, according to some arbitrary cut-off point, and that progression through grades is automatic.

If these assumptions were valid, the age and schooling effects are estimated by means of a regression discontinuity design (Cook & Campbell, 1979), involving regressions of test scores on chronological age. The effect of age is reflected by the slope of the within-grade regressions, whereas

the effect of schooling is reflected in the discontinuity between the two regression lines.

The first assumption of the model is reasonably met. The second is more problematic because, in practice, the admission to school is not solely a matter of the child birth date. As mentioned in the introduction, relatively bright children might enter the first grade "early," whereas children who are not sufficiently developed (intellectually or emotionally) remain an additional year in kindergarten. The frequency of grade misplacement is particularly high near the official cut off point (which in Israel is based on the Hebrew calendar and falls sometime in December; see Cahan & Cohen, 1989 for details). In order to cope with this problem of selection, we excluded from the computation of the within-grade regressions two groups of children: (1) children who did not fall into the official age range of their cohort and (2) first graders born in November or December 1982 (i.e., the oldest in their class), the months with the highest proportion of "missing" children (Cahan & Cohen, 1989).

Subjects. The sample consisted of all first graders born in 1981 (with the exceptions described above) frequenting the seven elementary schools serving four neighborhoods of Jerusalem (319 children of both genders), and all children born in 1982 from the 19 kindergartens serving the same neighborhoods (352 children of both genders). The selected neighborhoods represented upper middle-class, middle-class, and lower-middle class population.

Tests and Materials. Phonological awareness was measured by a battery of four sub-tests of constrained phonemic segmentation (Goldstein, 1976; Zhurova, 1973) each containing 20 items. The sub-tests were selected from a battery devised and validated in a pilot study (H. Leshem, unpublished doctoral dissertation), and were chosen because they did not require subjects to perform cognitive operations other than phonemic segmentation (for a survey of various types of segmentation tests see Content, Kolinsky, Morais, & Bertelson, 1986; Stanovitch, Cunningham, & Cramer, 1984). The tasks were:

1. Isolation of the first phoneme in spoken words. The children were instructed to utter the first phoneme in words pronounced by the examiner.

2. Isolation of the first phoneme in self generated pictures' names. The children were shown pictures of common objects and asked to pronounce the first phoneme in the name of each object.

3. Isolation of the last phoneme in spoken words. Similar to test 1 except that the last phoneme had to be isolated. The words were different than in test 2.

4. Isolation of the last phoneme in self generated pictures' names. Similar to test 2 except that the last phoneme in the name of each object had to be isolated. The objects were different than in test 2.

The words and object names were selected in collaboration with teachers in the respective grades to be part of the children's vocabulary. They were uni- to three-syllabic words. Both consonants and vowels were used as initial or last phonemes.

Measures of phonological awareness. The phonological awareness score of each child was the percentage of correct responses across all four subtests. In addition, two error scores were calculated per subject: (1) The percentage of syllabic (rather than phonemic) segmentation. (2) The percentage of sub-syllabic (i.e., consonant + vowel) segmentation. This distinction was particularly desirable in this study because in Hebrew vowels are represented primarily by diacritical marks that are always appended to consonantal letters. Hence, the basic phonemic unit that is mostly emphasized by teachers during the processes of reading acquisition is bigger than a single phoneme, including a consonant and a vowel. In many cases, however, this CV unit does not form a syllable. Thus, it is possible that, unlike in Italian or English, in Hebrew learning to read should develop some awareness to sub-syllabic rather than phonemic segments.

Procedure. The entire sample was tested within the last two weeks of February. Hence, the school children had 5 months of reading instruction. The examiners were 20 students of education or psychology who received special training; they were sent at random to first grade classes and kindergartens and most tested both groups of children.

The tests, which lasted together from 30 to 40 minutes, were administrated individually in a separate room in the school (or kindergarten). Before performing each task, the child was given a fixed number of practice items, preceded by an example. During practice, but not during the test, feedback was provided and errors were corrected.

Results

As expected, the percentage of correct responses on the phonemic segmentation battery was higher in school children (76%, SD=14%), than in the

kindergarten (35%, SD=23%) ($t(674)=29.12$, $p<.0001$). This difference reflects the combined effects of age and schooling. The separate effects of these two factors are revealed in the analysis of the within-grade linear regressions of phonological awareness scores on age (Figure 1).

Owing to the insignificant difference in the slopes obtained within each grade level, it was assumed that the two regression lines were parallel. Accordingly, the net effects of chronological age and schooling were obtained from the regression coefficients of age (in months) and grade level in the multiple regression equation of test scores on age and grade. The net effect of one year difference in chronological age was 9% (SE=3.0%), and the net effect of one year of schooling was 32% (SE=3.4%) (see Figure 1). Both effects and the difference between them were significant ($p<.05$).

As would be expected, improved phonemic segmentation, whether as a function of chronological age or of schooling, was accompanied by a reduction in the percentage of errors. Separate analyses of the effects of schooling and age on syllabic and subsyllabic segmentation revealed that schooling had a larger effect than aging in reducing both types of errors. However, while schooling reduced syllabic segmentation more than CV segmentation, the effect of maturation was bigger on CV than on syllabic segmentation (Table 1).

Table 1. *Percentage (SD) of syllabic and sub-syllabic segmentation errors made by kindergarten and first grade children.*

	Kindergarten	Grade A
Syllabic errors	12 (5)	8 (6)
Sub-syllabic errors	27 (13)	13 (7)

Discussion

The results of the present study point to schooling as a major factor affecting the development of phonological awareness. While they prove that an age difference of one year significantly improves performance on some segmentation tests, the present results revealed that the experience accumulated during the first five months of schooling enhanced phonological awareness four times as much. This effect was impressive in both absolute and relative terms: 32% correct answers corresponds to an effect size of 1.4 kindergarten standard deviations, which is an unusually large effect.

PHONOLOGICAL AWARENESS IN CHILDREN

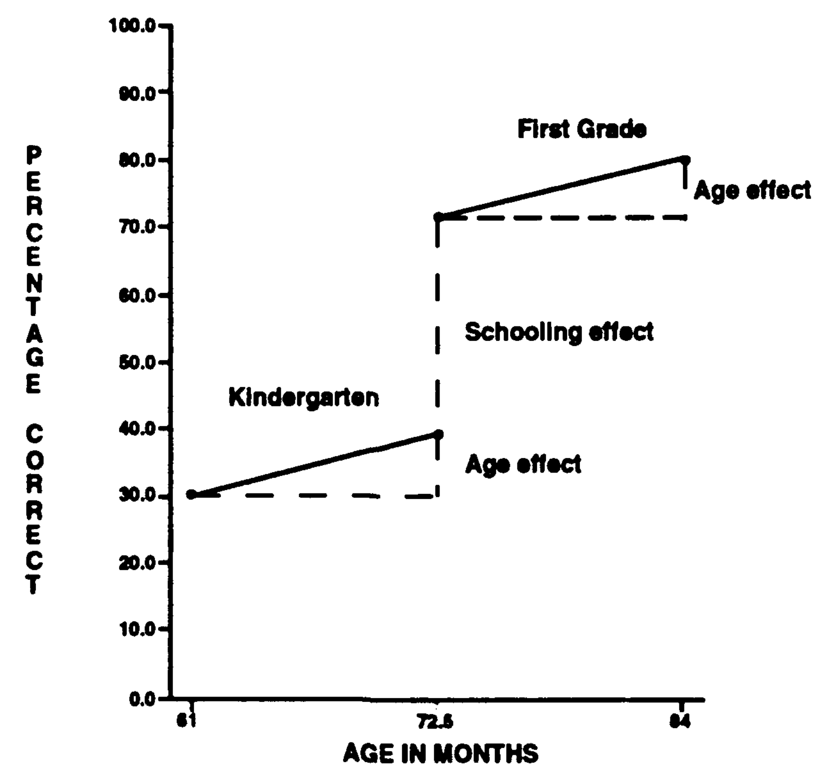

Figure 1. **The regression of phonological awareness scores on age in kindergarten and school (Grade I) children.**

Interpreting the schooling effect we should consider that we tested our sample during the last two weeks of February. Hence, this effect is based on only the first five months in school. Although during the first grade Israeli school children are involved in a variety of scholastic topics, the main curricular activity during the first half of the year is dedicated almost entirely to reading instruction. At the same time, the kindergarten activity includes no formal exposure to the alphabet. Consequently, we suggest that the schooling effect reflects primarily reading instruction and, therefore, that the present results support the contention that learning to read significantly enhances phonological awareness.

Additional support for a connection between reading instruction and the development of phonological awareness is provided by the analysis of errors. Indeed, the method of reading instruction adopted by a great majority of Israeli schools ("without secrets") emphasizes the sound of individual orthographic segments. However, as already mentioned, many orthographic segments in Hebrew are, in fact, mapped into two phonemes, a consonant and a vowel. Accordingly, although schooling reduced errors caused by sub-syllabic (CV) as well as syllabic segmentation, the former were reduced less. This trend contrasts the usual findings in other languages where a direct transition from syllabic to phonemic segmentation was observed (e.g. Cossu et al., 1988), and is best explained by the specificity of the Hebrew orthography. Thus, the schooling effect on the pattern of errors suggests that reading instruction foster phonological awareness by manipulating language-specific orthographic segments. The latter hypothesis was supported by the results of a recent study of bilingual children (Bentin & Bork, unpublished). The results of that study showed that learning to read Hebrew improved performance on segmentation tests in English only about half as much as in Hebrew.

The significant influence of the process of reading acquisition on the development of phonological awareness should not, however, be interpreted as evidence against the importance of phonological awareness on reading acquisition. In fact, several studies revealed that improving phonological skills in kindergarten has a positive influence on reading acquisition (Bradley, 1989; Bradley & Bryant, 1983; 1985, Bentin & Leshem, in press; see also Perfetti, Back, Bell, & Hughes, 1987; Vellutino & Scanlon, 1987; for a recent review see Goswami & Bryant, 1990). Moreover, the significant age effect that was observed in the present study suggests that some forms of phonological awareness is achieved in kindergarten and is independent of formal reading instruction.

These data suggest that cognitive-linguistic skills that are necessary for achieving phonological awareness mature by the age of six, promoted by natural development and/or informal linguistic experience. It is possible that this maturation is a necessary condition for reading acquisition in the first grade to trigger phonological awareness.

The significant within-grade (age) effect is more difficult to interpret. Obviously, this effect can be due to spontaneous cognitive maturation. However, maturation is not the only possible explanation. Six years old children are not only one year older than five years old children but also more experienced in areas that might be relevant to phonological awareness. Although in Israel formal instruction in the kindergarten does not include learning the alphabet, the children are informally exposed to orthographic symbols while watching TV, street signs, etc. The amount of informal experience with letters is proportional to age. Therefore, the within-grade increase in phonological awareness observed in the present study might reflect the increased linguistic experience rather than "pure" cognitive maturation. In other words, both the "grade level" and the "age level" effects in the present study might have been mediated by the same underlying factor, the amount of experience with printed language. Hence, the difference between the two effects might reflect the difference between formal reading instruction and informal experience with printed language.

Before concluding, one caveat should be considered. In the present study, we tested phonological awareness by tests of phonemic segmentation. Other studies suggest that the present results might not be valid for other tests of phonological awareness. For example, syllabic segmentation ability was quite good in kindergarten (Bentin & Leshem, in press, Liberman et al., 1974) and that sensitivity to rhymes and alliterations develops naturally between the age three and five, before the children can read (Maclean, Bryant, & Bradley, 1987). Different effects of literacy on phonemic and syllabic or sub-syllabic segmentation was found also in illiterate adults (Bertelson & de Gelder, 1989; Bertelson, de Gelder, Tfouni & Morais, 1989). That study showed the illiterates performed reasonably well in tests of vowel

deletion and rhyme judgment, but poorly on consonant deletion. On the basis of their findings, Bertelson et al., (1989) propose that phonological awareness is a heterogenous meta-linguistic ability that involves "involve separate components which obey different developmental mechanisms." Considering the existing pattern of evidence including our own, we adhere to this proposition. We suggest that sensitivity to highly resonant vocalic centers that form syllabic nuclei develops naturally during speech perception. On the other hand, explicit deciphering of coarticulated individual phonemes and ability to consciously manipulate phonemic segments is significantly enhanced by learning to read an alphabetic orthography.

REFERENCES

Alegria, J., Pignot, E., & Morais, J., (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory and Cognition, 10*, 451-456.

Bentin, S., & Bork, H. L (in preparation). Inter-languages transfer of training phonological awareness: A study of schooling effects on the development of phonological awareness in Hebrew English bilinguals.

Bentin, S., & Leshem, H. (in press). On the interaction of phonological awareness and reading acquisition. It's a two-way street. *Annals of Dyslexia*.

Bertelson, P. (1987). The onset of literacy: Cognitive processes in reading acquisition. Cambridge, MA: MIT Press.

Bertelson, P., de Gelder, B. (1989). Learning about reading from illiterates. In A. M. Galaburda (Ed.), *From reading to neurons* (pp. 1-26). Cambridge Massachusetts: MIT Press.

Bertelson P., de Gelder B., Tfouni L. V., & Morais J. (1989). Metaphonological abilities of adult illiterates: New evidence of heterogeneity. *European Journal of Cognitive Psychology, 3*, 239-250.

Bertelson, P., Morais, J., Alegria, J., & Content, A. (1985). Phonetic analysis capacity and learning to read. *Nature, 313*, 73-74.

Bertelson, P., Morais, J., Cary, L., & Alegria J. (1987). Interpreting data from illiterates: Reply to Koopmans. *Cognition, 27*, 113-115.

Bradley, L. (1989). Predicting learning disabilities. In J. J. Dumont & H. Nakken (Eds.) *Learning disabilities: Cognitive, social and remedial aspects* (pp. 1-17). Amsterdam: Swets & Zeitlinger.

Bradley, L., & Bryant, P. B. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419-421.

Bradley, L., & Bryant, P. E. (1985). Rhyme and reason in reading and spelling, IARLD Monographs No. 1. Ann Arbor: University of Michigan Press.

Cahan, S., & Cohen, N. (1989). Age vs. schooling effects on intelligence development. *Child Development, 60*, 1236-1249.

Cahan, S., & Davis, D., (1987). A between-grades-level approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal, 24*, 1-13.

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in pre-readers: Effects of corrective information. *Journal of Experimental Child Psychology, 42*, 49-72.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design analysis issues for field settings.* Boston: Rand McNally.

Cossu, G., Shankweiler, D., Liberman. I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1-16.

Fox, B., & Routh, D. (1975). Analyzing spoken language into words, syllables and phonemes: A developmental studies. *Journal of Psycholinguistics Research, 4*, 331-342.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Educational Psychology, 68*, 680-688.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Lawrence Erlbaum Ltd.

Liberman, A. M., & Mattingly, I. G. (1989). *Science, 243*, 489-494

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

Liberman, I. Y., Shankweiler, D. Fischer, F. W., & Carter, B. J. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201212.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

de Manrique, A. M. B., & Gramigna, S. (1984). La segmentacion Fonologica y silabica en ninos de preescolar y primer grado. *Lectura y Vida, 5*, 4-13.

Maclean, M., Bryant, P., & Bradley, L, (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly, 33*, 255-281.

Morais, J., Bertelson, P., Cary, L., & Alegria, J., (1986). Literacy training and speech segmentation. *Cognition, 74*, 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Morais, J., Castro, M. S. L., Scliar-Cabral, L., Kolinsky, R., & Content, A. (1987). The effects of literacy on the recognition of dichotic words. *Quarterly Journal of Experimental Psychology, 39A*, 451-465.

Morrison, F. J. (1988). Development of phonemic awareness: A natural experiment. Paper presented at the Annual Meeting of the Psychonomic Society, Chicago.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.

Read, C., Zhang, Y., Nie, H., & Ding B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic spelling. *Cognition, 24*, 31-44.

Stanovitch, K. E., Cunningham, A. E., & Crammer, B. R. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*, 175-190.

Treiman, R. A., & Baron, J. (1981). Segmental analysis: Development and relation to reading ability. In G. E., Mackinnon, & T. G. Walker (Eds.), *Reading research: Advances in theory and practice*, 3 (pp. 159-198) New York: Academic Press.

Vellutino, F. R . & Scanlon, D. (1987) *Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. Merrill-Palmer Quarterly*, 33, 321-363.

Zhurova, L. Y. (1973). The development of analysis of words into their sounds by preschool children. In C. A. Ferguson, & D. I. Slobin (Eds.), *Studies of child language development* (pp. 141154), New York: Holt, Rinehart, & Winston.

FOOTNOTES

**Psychological Science, 2(IV)*, 271-274 (1991).

†Department of Psychology and School of Education, The Hebrew University, Jerusalem.

††School of Education, The Hebrew University, Jerusalem.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 147-162

Bi-alphabetism and the Design of a Reading Mechanism*

Laurie Beth Feldman†

Evidence for alphabetically-defined visual effects was examined in six word recognition studies with Serbo-Croatian materials. In each, the experimental manipulation exploited the bi-alphabetic fluency of skilled adult readers and compared performance on a variety of measures using successive presentations of words and pseudowords in the same or in different alphabets. One line of investigation manipulated number of intervening items in a repetition priming version of the lexical decision task. A second line of investigation used alphabet decision as a study phase prior to lexical decision. A third examined lexical decision and naming latencies to targets in phonologically and graphemically similar and dissimilar (prime) contexts. In none of these studies did alternating as contrasted with preserving alphabet exert a significant effect on word recognition. Three additional related lines of inquiry examined the effect of alphabetic context on words that are phonologically ambiguous because they can be interpreted as either Roman or Cyrillic letter strings and on words that are phonologically unambiguous because they can be interpreted in only one way. Alphabetic context influenced the processing of phonologically ambiguous words but not of unambiguous words both when the availability of the context was restricted, either in its duration or by the presence of a pattern mask, and when it was not. It was concluded that, alphabetically-defined visual effects in Serbo-Croatian word recognition reveal themselves under conditions of phonological complexity. Results are described in terms of a connectionist model with letter-, phoneme- and word-sized units where alphabetic effects arise in the mapping between letter and phoneme levels.

The linguistic conditions in regions of Yugoslavia provide an ideal medium in which to investigate the role of a word's visual form in the process of word recognition. Specifically, two visually distinct alphabets, Roman and Cyrillic, are used interchangeably and with impressive fluency by most skilled readers in the Belgrade region. Consequently, words of Serbo-Croatian, the official language of Yugoslavia, can be written in either the Roman or the Cyrillic alphabets and, according to the educational policy in effect until recently, all school children are required to demonstrate and maintain proficiency in both alphabets. The implication of the forgoing is that skilled readers of Serbo-Croatian maintain two visually-defined lexicons or at least, two visually-defined descriptions for each word. And, because most of the phonemes are unique to one alphabet or another, the visual similarity of the two alphabetic transcriptions of a word is dramatically reduced relative to the experimental manipulations of visual form (e.g., case) that are possible in English. In addition, the writing system for Serbo-Croatian was reformed in the last century so that the mapping of letter to sound is consistent and regular. The implication of a phonologically-regular writing system is that skilled readers of Serbo-Croatian need never rely on word-level knowledge in order to arrive at the correct phonemic form of a word.

The present chapter summarizes six lines of investigation using variations on the lexical decision and naming methodologies that were conducted with bi-alphabetically fluent readers of Serbo-Croatian (see Table 1). Collectively, they investigate the role of an alphabetically-defined (visual) level of description in word recognition.

This research was supported by grant HD 01994 to Haskins Laboratories and grant HD 08495 to the University of Belgrade from the National Institutes of Child Health and Human Development. Results of Study 1 were presented to the 1989 meeting of the Psychonomic Society in Atlanta, GA.

Table 1. *Summary of Lexical Decision (LD), Alphabet Decision (AD), and Naming (N) studies.*

Study	experimental conditions	first item	example	second item	example	reference
1	alph alternated	LD	ногом (C)	LD	NOGA (R)	Feldman, unpublished
	alph preserved	LD	NOGOM (R)	LD	NOGA (R)	
2	alph alternated	AD	ногA (C)	LD	NOGA	Feldman, unpublished
	alph preserved	AD	NOGA (R)	LD	NOGA	
3	phon matched					Lukatela & Turvey, 1990a
	alph matched	No response	PASUS (R)	LD	RACUN (R)	Lukatela, Carello & Turvey, 1990
	phon matched					
	alph mismatched	No response	ПАСУС (C)	LD	RACUN (R)	
	phon mismatched					
	alph matched	No response	DUVAČ(R)	LD	RACUN (R)	
	phon mismatched					
	alph mismatched	No response	ЛУВАЧ(C)	LD	RACUN (R)	
4	phon ambiguous	None		LD/N	BEHA	Feldman & Turvey, 1983; Feldman, Kostić,
	phon unambiguou	None		LD/N	VENA	Lukatela & Turvey, 1983; Feldman, 1991
5	sem. associated	No response	ОЛУJA	LD	BETAP (C)	Lukatela, Feldman, Turvey, Carello & Katz, 1989
	alph. consistent	No response	"storm" (C)		"wind"	Lukatela, Turvey, Feldman, Carello & Katz, 1989
	sem associated	No response	OLUJA	LD	BETAP (C)	
	alph. inconsistent	No response	storm" (R)			
	sem unassociated	No response	ПАС	"LD	BETAP (C)	
	alph consistent	No response	dog" (C)			
	sem unassociated	No response	PAS	"LD	BETAP (C)	
	alph inconsistent	No response	dog" (R)			
	sem associated	No response	OLUJA	"LD	VETAR (R)	
	alph consistent	No response	storm" (R)		"wind"	
	sem associated	No response	ОЛУJA	"LD	VETAR (R)	
	alph inconsistent	No response	storm" (C)			
	sem unassociated	No response	PAS	"LD	VETAR (R)	
	alph consistent	No response	dog" (R)			
	sem unassociated	No response	ПАС	"LD	VETAR (R)	
	alph inconsistent	No response	"dog" (C)			
6	alph consistent	No response	ЛДГ (C)	LD/N	BETAP (C)	Lukatela, Turvey & Todorović, 1991
	letter(s)	No response				Lukatela & Turvey, 1990b
	alph inconsistent	No response	LDG (R)	LD/N	BETAP (C)	
	letter(s)	No response				
	alph consistent	No response	LDG (R)	LD/N	VETAR (R)	
	letter(s)	No response				
	alph inconsistent	No response	ЛДГ (C)	LD/N	VETAR (R)	
	letter(s)	No response				

All of the studies exploit the particular relation between two alphabets that exists in Yugoslavia and all were conducted with first year students at the University of Belgrade or with advanced high school students in the Belgrade region who are fluent in both alphabets. Studies One and Two focus on the visual distinctiveness of orthographic forms. Specifically, most phonemes of Serbo-Croatian have two quite distinct visual forms, one Roman character and one Cyrillic character, and this variation provides a tool with which to ask whether multiple presentations that preserve alphabetically-defined visual patterns facilitate performance relative to presentations that alternate alphabet. Study three examines facilitation due to visual and phonological similarity for words presented close in succession. The remaining three studies exploit properties of the subset of characters that are shared by the two alphabets. Specifically, there are a small number of phonemes where the mapping between letter and phoneme is complex because the same visual characters are shared by both alphabets. Of these shared characters, the *common* characters (i.e., A, E, O, J, K, M, T) receive the same phonemic interpretation in both alphabets whereas the *ambiguous* characters (i.e., B, C, H, P) represent different phonemes in Cyrillic and in Roman (see Table 2). Comparisons between words composed exclusively of shared characters (i.e., words with two phonemic interpretations) and words that include at least one nonshared (i.e., alphabetically unique) character provide the basis of studies four, five, and six where the effect of alphabetic context on phonological processing is explored. To anticipate, this chapter will review a series of studies that explores the graphemic and phonemic implications of reading in two alphabets and will provide a model of word reading in Serbo-Croatian with its emphasis on phonology. Because the first two studies are not published and details are not easily obtained, they will be described in more detail than will subsequent studies.

Study 1: Alphabetic manipulations across repetitions of a word

One way in which the bi-alphabetic fluency of readers of Serbo-Croatian has been exploited has been to investigate the role of alphabetically-defined orthographic similarity of prime and target in repetition priming (Feldman & Moskovljević, 1987, Expt. 1). In this task, words and pseudowords are presented twice, with a lag of intervening items, and subjects are instructed to perform a lexical decision to each letter string as it appears (Stanners, Neiser, Hernon & Hall, 1979). The critical experimental manipulation entailed repetitions in either the same or in different alphabets. In the *alphabet alternated* condition, prime and target were transcribed in different alphabets (e.g., NOGOM-NOGOM). In the *alphabet preserved* condition, prime and target were in the same alphabet (e.g., NOGOM-NOGOM). Equal numbers of words and pseudowords were presented for durations of 750 ms. The interval between successive presentations of a word averaged 10 items with a range of 7 to 13. One group of subjects saw all items in Roman script (alphabet preserved) and the other saw primes in Cyrillic and targets in Roman (alphabet alternated). Results indicated that facilitation (i.e., reaction time to first minus second presentation) was numerically equivalent (viz., 90 ms) in the alphabet preserved and the alphabet alternated conditions. The authors interpreted this pattern of results as evidence that at lags of 7 to 13, visual similarity of prime and target alone did not provide a source of facilitation in the repetition priming task.

Because it is possible that the time course of activation of visual form varies with lag (Monsell, 1985; Ratcliff, Hockley, & McKoon, 1985), the first study attempted to replicate this finding. In addition, consistency of alphabet was systematically manipulated. Decision latencies to targets that were preceded by primes (where target and prime either alternated or preserved alphabet) were compared over lags of 10 and 20 (Experiment 1a) or lags of 3 and 10 (Experiment 1b) in an attempt to find evidence for facilitation based on repetitions of specific visual patterns. Materials consisted of thirty two Serbo-Croatian words and thirty two pseudowords. Words were familiar nouns in nominative case that contained three or four letters. Pseudowords were generated by changing one or two letters (vowel with vowel or consonant with consonant) and preserved orthographic and phonemic regularity.

Each word and pseudoword appeared two times, once as a target and once as a prime and, as noted above, the lag or interval between presentation of prime and its target was varied. Half of the targets were printed in upper case Roman and half were printed in upper case Cyrillic. And, at each lag, half of the prime-target pairs alternated alphabet and half preserved it. Items were selected so that both alphabet transcriptions included at least one letter that uniquely specified alphabet (Feldman, Kostić, Lukatela, & Turvey, 1983).

Table 2. *Letters unique to the Roman and/or Cyrillic alphabets and letters shared by the Roman and Cyrillic alphabets.*

Roman grapheme	Roman phoneme	Cyrillic phoneme	Classification	Cyrillic grapheme	Cyrillic phoneme	Roman phoneme	Classification
A	/a/	/a/	common[1]	А	/a/	/a/	common
B	/b/	/v/	ambiguous[2]	Б	/b/		Cyrillic
C	/ts/	/s/	ambiguous	Ц	/ts/		Cyrillic
Ć	/tʃj /		Roman	Ч	/tʃ/		Cyrillic
Č	/tʃ/		Roman	Ћ	/tʃj/		Cyrillic
D	/d/		Roman	Д	/d/		Cyrillic
Đ	/dʒj/		Roman	Ђ	/dʒj/		Cyrillic
DŽ	/dʒ/		Roman	Џ	/dʒ/		Cyrillic
E	/e/	/e/	common	Е	/e/	/e/	common
F	/f/		Roman	Ф	/f/		Cyrillic
G	/g/		Roman	Г	/g/		Cyrillic
H	/x/	/n/	ambiguous	Х	/x/		Cyrillic
I	/i/		Roman	И	/i/		Cyrillic
J	/j/	/j/	common	Ј	/j/	/j/	common
K	/k/	/k/	common	К	/k/	/k/	common
L	/l/		Roman	Л	/l/		Cyrillic
LJ	/lj/		Roman	Љ	/lj/		Cyrillic
M	/m/	/m/	common	М	/m/	/m/	common
N	/n/		Roman	Н	/n/	/x/	ambiguous
NJ	/nj/		Roman	Њ	/nj/		Cyrillic
O	/o/	/o/	common	О	/o/	/o/	common
P	/p/	/r/	ambiguous	П	/p/		Cyrillic
R	/r/		Roman	Р	/r/	/p/	ambiguous
S	/s/		Roman	С	/s/	/ts/	ambiguous
Š	/ʃ/		Roman	Ш	/ʃ/		Cyrillic
T	/t/	/t/	common	Т	/t/	/t/	common
U	/u/		Roman	У	/u/		Cyrillic
V	/v/		Roman	В	/v/	/b/	ambiguous
Z	/z/		Roman	З	/z/		Cyrillic
Ž	/ʒ/		Roman	Ж	/ʒ/		Cyrillic

[1]Common letters have the same interpretation in Roman and Cyrillic. [2]Ambiguous letters have different interpretations.

A small number of filler items were introduced to maintain the appropriate lags. Across test orders each target (word or pseudoword) was preceded by its prime at two different lags in both the alphabet alternated and alphabet preserved conditions. University students were tested individually in a lexical decision task. Each subject viewed one test order and a practice list of ten items preceded the test list.

In addition to eliminating errors and extreme response times, responses were excluded when a subject responded incorrectly to one member of a prime-target pair. Table 3 summarizes the mean recognition times over subjects for target words and pseudowords as a function of lag for alphabetically alternated and preserved pairs.

Analyses of variance on targets from Experiment 1a with lag (10, 20) and alphabet (alternated, preserved) as independent variables were performed separately for words and pseudowords using subjects (*F1*) and items (*F2*) as random variables. For pseudowords, no effects or interactions were significant. For words, the effect of alphabet was marginally significant in the analysis of latencies by subjects $F1(1,35)= 4.08$, $MSe= 967$, $p <.051$ but did not approach significance in the analysis by items. Neither the effect of lag nor the interaction of alphabet by lag was significant. Similarly with errors, no main effects or interactions approached significance. Finally, the pattern observed with errors did not support the latency pattern.

When lags of 3 or 10 items separated prime and target, analyses performed on latencies for target items alone revealed no significant effect of alphabet, no effect of lag and no interaction. Analogously, the error scores were not sensitive to manipulations of lag or alphabet. The interaction of lag by alphabet was not significant for pseudowords.

The present study exploited the bi-alphabetic knowledge of Yugoslav readers in order to investigate the role of visually-defined similarity as a source of facilitation in the repetition priming paradigm. In contrast to the design used in Feldman and Moskovljević (1987), the present design treats alphabet consistency of prime and target as a within-subjects variable. Two experiments were conducted and, across experiments, the average lag was manipulated. In neither experiment was the effect of lag significant for words. Neither at a lag of 3 nor at a lag of 20 did facilitation differ significantly from lag 10.

Table 3. *Mean decision latencies (ms) and errors for words and pseudowords in the alphabet preserved and alphabet alternated conditions of the repetition priming task for Study 1.*

	First	Repetition Alphabet			
	Presentation	Lag	Alternated	Preserved	Difference
Experiment 1a					
words	651	10	601	592	9
	12.7		6.6	7.3	-0.7
		20	607	595	12
			4.5	7.3	-2.8
pseudowords	666	10	682	680	2
		5.7	5.2	5.9	-0.7
		20	672	661	11
			4.2	5.9	-1.7
Experiment 1b					
words	628	3	562	562	0
		10.8	8.3	5.9	2.4
		10	567	573	-6
			7.9	7.9	0
pseudowords	654	3	672	665	7
	5.7	6.9	7.6		-0.7
		10	648	629	19
			6.1	6.9	-0.8

The main finding was that for both words and pseudowords, significant target facilitation occurred when primes appear in either the same alphabet or in a different alphabe· ·rom the target. Importantly, target facilitatioı. was no greater in the alphabet preserved condition than in the alternating condition. Small numerical differences that were sometimes observed with the latency measure were not supported bv the error measure. The intent of the alphab· ecision study was to demonstrate an effe. : prior experience with specific visual forms of words and pseudowords on subsequent lexical decision performance with those same materials. The experiment exploited a special characteristic of Serbo-Croatian, notably the multiple mapping from phoneme to graphemes that exist because readers are fluent in both the Roman and Cyrillic alphabets. Facilitation defined either in terms of the difference between first and second presentations or as a percent decrease in lexical decision latency (relative to the first presentation) were not significantly different for alphabet preserved and alphabet alternating conditions.

Words presented and represented in the same alphabet are more visually similar than are the Roman and Cyrillic transcriptions of a word. Yet, in the repetition priming task where several items intervened between first and second presentations, no significant increment to facilitation was observed on the alphabet preserved trials relative to the alphabet alternating trials. This outcome is not surprising if, as Masson and Freedman (1990) have claimed, visual analysis (e.g., improved perceptual sensitivity) is not responsible for the repetition effect (p. 356) but rather, the bases of facilitation for repeated items are more conceptual interpretive processes that are associated with a shift in decision bias. Perhaps, because of the nature of the experimental task, an analysis of the alphabet manipulation within a repetition priming task cannot provide compelling evidence for the role of visual analysis and orthographic representations in word recognition.

Study 2: Alphabetic manipulations in a alphabet decision task

The pattern of facilitation in the repetition priming task with a within-subjects manipulation of alphabet provided no evidence that, in the course of visual word recognition, subjects are constrained by an orthographic representation based on the visual form of the letter string. Although the previous task did not foster a visual analysis, it is plausible that skilled readers of Serbo-Croatian who are fluent in two alphabets can. under the proper circumstances, engage in an ana··sis of a letter string that retains its visual characteristics and this is the focus of the second study. In the first phase of study 2, subjects were told to attend to the alphabetic characteristics of the letter stings that they encountered. They were instructed to indicate the alphabet in which each letter string was printed by a key press. In a second phase, they were asked to make a lexical decision to those same letter strings. The goal was to try to induce subjects to attend to the visual attributes of the materials that they encountered in an attempt to demonstrate that skilled readers of Serbo-Croatian can attend to the visual characteristics of a letter string.

Forty-four first year students from the Department of Psychology at the University of Belgrade participated in the experiment. Half of the subjects part··ated in an alphabet decision task and then in .. lexical decision task. The remaining half participated only in the lexical decision task. Experimental targets consisted of forty Serbo-Croatian words and forty pseudowords. Words were familiar nouns in nominative case tha· ··ntained three or four letters. As in the prev· ·· study, pseudowords were generated by changing one or two letters (vowel with vowel or consonant with consonant) and preserved orthographic and phonemic regularity. In both the alphabet decision and the lexical decision phases of study 2, half of the words and half of the pseudowords were printed in Roman and half were printed in Cyrillic. Items were selected so that both alphabet transcriptions included at least one letter that uniquely specified alphabet.

As each letter string appeared on the CRT of an Apple II in the alphabet decision task, subjects pressed either of two telegraph keys with both hands to indicate alphabet. In the second phase of the experiment, the same words and pseudowords were presented to subjects in a different order. Subjects performed a lexical decision to each letter string. The presentation format was identical to the alphabet decision phase described above. Reaction time was measured from the onset of the letter string.

In the lexical decision phase, as in the alphabet decision phase, half of the ite·· were in Cyrillic and half were in Romar ··d words and pseudowords were equally represented in each alphabet. In the lexical decision phase, however, half of the words and half of the pseudowords preserved the alphabet of their earlier

presentation and half alternated alphabet. In this study, alphabet (preserved or alternated) and lexicality (word or pseudoword) were manipulated within subjects and prior participation in the alphabet decision task was manipulated between subjects. Results revealed a significant effect of prior alphabet decision on performance in the lexical decision task. Subjects who participated in lexical decision following alphabet decision were significantly slower than subjects who participated only in the lexical decision task. This outcome is consistent with the observation that repetition effects are sensitive to the task at initial presentation and do not always reveal themselves as facilitation (Forster & Davies, 1984; Ratcliff et al., 1985; Bentin & Peled, 1990). Subsequent analyses were conducted on the lexical decision following alphabet decision data.

Mean latencies and error scores for the lexical decision phase are summarized in Table 4. (Scores greater than 1200 ms or less than 400 ms were treated as errors and eliminated from the reaction time analyses.) An analysis of variance on latencies revealed a significant effect of lexicality $F1(1,21)= 14.98$, MSe= 1110, $p <.001$; $F2(1,78)= 10.16$, MSe= 3147, $p <.003$. Neither the effect of alphabet nor the interaction of lexicality by alphabet approached significance. No effects were significant with errors as the dependent measure and the small numerical differences diverged in direction from the small latency differences.

Table 4. *Mean decision latencies (ms) and errors for words and pseudowords in the lexical decision phase of the alphabet decision task.*

	Alternated	Alphabet Preserved	Difference
words	712	702	10
	3.9	4.6	0.7
pseudowords	737	732	5
	2.3	3.9	-1.6

The intent of the alphabet decision study was to demonstrate an effect of prior experience with specific visual forms of words and pseudowords on subsequent lexical decision performance with those same materials. By using both Roman and Cyrillic characters, orthographic form was either preserved or alternated across the alphabet and lexical decision phases of the study. The logic of the first phase of the study was to direct subjects to attend to alphabet and their accuracy levels proved that they could do this. The effect of attending to alphabet on subsequent word recognition was then examined.

Relative to performing a word level task in isolation, subjects were slower when they performed a letter level task such as alphabet decision prior to performing a word level task. The analysis of decision latencies in the second phase revealed a significant effect of lexicality on decision latency but no effect of alphabet. With respect to visual effects, viewing a word or a pseudoword twice in the same visual form (alphabet preserved) exerted no effect over and above the effect of viewing a word (or a pseudoword) once in its Roman form and once in its Cyrillic form (alphabet alternated). Moreover, the small numerical differences that were observed with the latency measure for the factor of alphabet were not supported by the accuracy measure. It appears that for recognition tasks at the level of the word, skilled readers of Serbo-Croatian, who tend to be equally fluent in both alphabets (Feldman & Moskovljević, 1987 footnote 1), cannot benefit from multiple presentations of alphabet-specific orthographic forms.

In a repetition priming task (study 1) and in an alphabet decision task that explicitly directed skilled readers to attend to alphabet (study 2), no effects of orthographic repetition were observed. While this is a null effect and it is possible that another task will be developed in which effects of alphabet-specific orthographic form can be demonstrated, it is evident that in two quite different word recognition tasks skilled readers do not appear to rely on a style of analysis that is primarily tied to the visual form of a word.

Study 3: Manipulations on alphabetic and phonemic similarity

It is plausible that the experimental conditions in the first two studies where repetitions were separated by a number of intervening items could not reveal effects of preserving or alternating because the interval between successive presentations exceeded the duration over which alphabet effects can persist. Alternatively, or conjointly, it is possible that no alphabetic effects were evident because all target items included at least one letter that uniquely specifies alphabet and alphabet effects emerge only when alphabet context is not well-specified. Accordingly, in a third line of investigation using a priming paradigm (Lukatela & Turvey 1990a), alphabetic effects at short lags are examined for target words that contain at least one unique letter.

In traditional priming paradigms, targets items are immediately preceded by a context or prime and in some experimental conditions, the context is related to the target along some dimension. Target latencies with and without related primes are compared. In contrast to the previous two studies where the first and second items were separated by other intervening items, in the third study, phonologically unambiguous targets were immediately preceded by contexts. Moreover, these contexts were related with respect to the dimensions of phonology, graphemic form, both or neither and subjects performed either a naming or a lexical decision task. Primes and targets were displayed serially, one immediately after the other in a presentation format that was likely to enhance similarity effects between prime and target.

First item (prime) and second item (target) consisted of either words or pseudowords. Items were phonemically matched or mismatched and were visually similar (alphabet preserved) or dissimilar (alphabet alternated). Primes appeared above the position of targets and disappeared 100 ms before the target was presented. Effects of phonological similarity were significant but direction varied with task. Visually similar primes had the same effect on target latencies as did visually dissimilar pairs in both the phonologically matched and the nonmatched conditions. For example, when primes and word targets differed in their initial phoneme and rhymed (i.e., phonologically similar condition), the difference between alternated (e.g., PAKUH-RAĈUN) and preserved alphabet (e.g., RAKUN-RAĈUN) latencies was 13 ms (0.22%) in lexical decision (Experiment 1; Lukatela & Turvey, 1990a) and 6 ms (0.43%) in naming (Experiment 5; Lukatela & Turvey, 1990a). Similar effects were observed for pseudoword targets. Effects of alphabet in the phonologically unmatched conditions of those experiments were even smaller. Stated generally, in study 3, preservation or alternation of alphabet was used as a manipulation of visual similarity and no effect of alphabetic similarity was observed for target letter strings that, because of the presence of at least one unique letter, were well-specified with respect to alphabet. Under sequential presentation conditions at inter-stimulus intervals of 100 ms there was no effect of graphemic similarity over and above the effect of phonological similarity.

The present result contrasts to analogous experiments conducted with English materials where phonemic similarity effects are difficult to obtain (compare Martin & Jensen, 1988 with Hillinger, 1980 and Meyer, Schvaneveldt & Ruddy, 1974, for example). With Serbo-Croatian materials, a robust effect of phonemic similarity was observed in the lexical decision task. Moreover, the direction of this effect depended on the position of the nonmatched letter and on the relative frequency of the context and target word. Relative to a phonologically dissimilar context, target-context pairs that differed in their initial letter showed facilitation (+55 ms) whereas pairs that differed on a medial letter showed slowing (-27 ms) (Experiment 2; Lukatela & Turvey, 1990a). Pairs with low target familiarity (uncommon words and pseudoword targets) showed facilitation (+51 ms) whereas high familiarity (word) targets showed slowing (-21 ms) (Experiments 3 and 4; Lukatela & Turvey, 1990a).

In the naming task, in contrast to the lexical decision task, facilitation due to phonological similarity was observed for both words and pseudowords with both initial and medial letter differences between context word and target. As in the lexical decision task, alphabetically-defined visual effects were never significant. Target familiarity had no effect (Experiments 5 and 6; Lukatela & Turvey, 1990a) although in naming, differences in word stress between context and target eliminated the effect of phonemic similarity (Experiment 9; Lukatela & Turvey, 1990a). When targets were highly familiar words and contexts were either real words or pseudowords, facilitatory effects of phonological similarity were observed in naming for both word and pseudoword contexts (Experiment 1; Lukatela, Carello & Turvey, 1990). In lexical decision, by contrast, phonemically similar word contexts produced inhibition while phonemically similar pseudoword contexts produced facilitation relative to dissimilar pairs (Experiment 2; Lukatela, Carello, & Turvey, 1990).

The effects of phonemic similarity of context and target were modelled as a network of letter, phoneme and word units such that constraints on the lexical decision task arise primarily at the level of word units that are partially activated by the phonemic units activated by the context. In the course of partially activating word units similar to the target (which generate inhibition to the target), phonemically similar contexts will also enhance the activation of the letter and phoneme units which comprise the target (Lukatela & Turvey, 1990a). In general, the dependence of context-target phonemic similarity on target familiarity in lexical decision reflects the balance

between inhibitory effects at the word level and excitatory effects at the letter and phoneme levels.

By contrast, the primary source of constraint on the naming task arises at the level of phonemic units which are sensitive to inputs from both the letter and word levels. The states and inhibitory relations among word units partially activated by the context are essentially irrelevant although, it is important to point out that naming a word benefits from activation of a word unit (and subsequent reinforcement of its phonemic constituents) in a way in which naming a pseudoword cannot. For letter strings that are well-specified with respect to alphabet, in both the lexical decision and naming tasks, effects of phonemic similarity arise from the use of phonological information activated by the context in the course of processing the target but no distinct influence of letter level (i.e., alphabetic) activation on phonological activity is evident.

Study 4: Alphabetic manipulations with phonological consequences

A fourth and very productive line of investigation into the consequences of two alphabetic systems probes the status, for the skilled reader, of words that are composed exclusively of letters that are shared by both alphabets. Results provide evidence of mandatory phonological processes prior to lexical access in word recognition tasks. As described in Table 2 (see also Turvey, Feldman, & Lukatela, 1980), some of these shared letters receive the same phonemic interpretation in both alphabets whereas others are phonemically ambiguous in that they receive different interpretations in Roman and in Cyrillic. Words composed exclusively of shared letters with the same phonemic interpretation in both alphabets (e.g., MAMA, JAJE) are alphabetically ambiguous but well-specified phonologically. Words composed of shared letters with two phonemic interpretations (and no alphabetically unique letters) are phonologically as well as alphabetically ambiguous in that they can be pronounced according to the grapheme-phoneme correspondence rules of Roman or those of Cyrillic or by combination of the two.

Consider the word BEHA which contains two phonologically ambiguous letters (viz., B, H) and two (alphabetically ambiguous but phonologically unique) common letters (viz., E, A). Interpreted as a Cyrillic letter string, it is pronounced /vena/ which means "vein." Interpreted as a Roman letter string, it is pronounced /bexa/ which is not a word in Serbo-Croatian, although it is a phonologically legal combination. A frequently replicated finding is that when skilled readers of Serbo-Croatian are presented with phonologically ambiguous letters strings in either the lexical decision or naming tasks, their responses are significantly slowed relative to their response latencies for phonologically unambiguous letter strings. In one study, (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978), both the design of the experiment and the instructions to the subjects were created to restrict the task to the Roman alphabet: No letter strings contained uniquely Cyrillic letters, and subjects were asked to judge whether a letter string was a word by its Roman reading. In a following study (Lukatela, Popadić, Ognjenović, & Turvey, 1980), no alphabet restriction was imposed on lexical decision and the word interpretation could occur in either the Roman interpretation, the Cyrillic interpretation, both or neither. In both experiments, the prolonged decision times to all phonologically bivalent letter strings as compared to phonologically unambiguous letter strings suggested that subjects are unable to suppress multiple phonological interpretations when permitted by a letter string. Because phonologically unambiguous letter strings with and without alphabet ambiguity produced equivalent results (e.g., MAMA which can be interpreted as either a Roman or a Cyrillic word was no slower than 'ABA which can only be interpreted as a Cyrillic string), this outcome was interpreted as evidence of phonological as contrasted with alphabetic ambiguity and it was concluded that lexical access always proceeds with reference to phonology.

A feature of the two experiments cited above (Lukatela et al., 1978; Lukatela et al., 1980) was that different words appeared in the phonologically unique and phonologically ambiguous conditions. That is, the effect of a letter string's phonological ambiguity was assessed by comparing recognition latencies of different words, some of which were phonologically ambiguous and some of which were not. Similarly, the effect of a letter string's alphabetic ambiguity was assessed by comparing recognition latencies of different (phonologically unambiguous) words, some of which were alphabetically ambiguous (e.g., MAMA, JAJE) and some of which were not (e.g., 'ABA, ŽABA)

In a later experiment, the effect of phonological ambiguity was assessed by comparing decision (Feldman & Turvey, 1983) and naming (Feldman, 1981) latencies to the ambiguous and unique tran-

scriptions of the same word. For example, the Serbo-Croatian word for "vein" is, as noted above, written BEHA in Cyrillic characters but, in Roman characters, that word is written VENA. Both forms are meaningful and are equated with respect to variables such as frequency, meaning and word length because they are forms of the same word. They differ, however, in that BEHA permits an alternative phonological interpretation (viz., /bexa/) whereas VENA does not. Comparisons between two alphabetic transcriptions of the same word, only one of which is phonologically ambiguous provide the basis of the within-word assessment of phonological complexity on word recognition in Serbo-Croatian known as the phonological ambiguity effect (PAE). Sometimes, differences as large as 300 ms have been observed between the ambiguous and unique alphabet transcriptions of a word although, among other factors, the magnitude of the PAE difference is sensitive to the number of ambiguous characters in the ambiguous form (Feldman, Kostić, Lukatela & Turvey, 1983; Feldman & Turvey, 1983). PAE effects have also been observed for ambiguous letter strings where neither (Feldman & Turvey, 1983) or both of the readings are meaningful (Frost, Feldman, & Katz, 1990).

To state the PAE outcome in a general way, prolonged latencies in naming and lexical decision have been observed for BEHA type words as contrasted with VENA type word but not for MAMA type words as contrasted with 'ABA or ŽABA type words. This outcome has been interpreted as reflecting activation of more phonemic units and competition among the word units to which they are linked (Feldman & Turvey, 1983; Feldman, Kostić, Lukatela & Turvey, 1983). A model, foreshadowed in the preceding discussion of phonemic and alphabetic similarity effects, has been proposed (Lukatela, Turvey, Feldman, Carello, & Katz, 1989). It consists of three types of units, letter, phoneme and word, and the linkages between them. At the level of the letter, the elements of the Cyrillic and Roman alphabets constitute functionally distinct sets. Shared letters with one phonemic interpretation (viz., A, E, O, J, K, M, T) are common to the two sets. Shared letters with two phonemic interpretation (viz., B, C, H, P) are represented in each alphabet set. That is, ambiguous letters are represented two times at the letter level.

At the level of the phoneme, by contrast, there is no duplication. Two grapheme units link to each phoneme unit (except for the shared letters that have the same phonemic interpretation in two alphabets). For example, F and F both connect to /f/ and B and V both connect to /v/ whereas A, which is both a Cyrillic and a Roman character, is the only unit that connects to /a/. The pattern of linkages between letter and phoneme units captures the relatively simple relation between letter and phoneme that characterizes the Serbo-Croatian language relative to a language such as English.

In the proposed network, word units are activated from phonemic units in a two-way interactive process. Each word unit represents a particular ordering of phonemic units. When a word unit is activated, the units at the letter and phoneme levels for each letter position in that word are reinforced. It is also assumed that there are multiple inhibitory connections (in both directions) between the unique letters of one alphabet and the unique letters of the other. So, for example, when a unique Cyrillic letter is activated in one position, then the activity level of all Roman letters in complementary positions is reduced. The strength of inhibition varies as a function of the number of activated units that are unique to one alphabet. In a similar manner, the strength and pattern of activation that gives rise to PAE varies as a function of the number of ambiguous units that are present (Feldman et al., 1983; Feldman & Turvey, 1983).

Consider a word such as BEHA which has phonemically ambiguous letters in the first and third positions. Each of these letters will activate two phonemic units (viz., B activates /b/ and /v/; H activates /x/ and /n/). Compare it with the Roman transcription of that same word, VENA, which has alphabetically unique letters in the first and third positions. The presence of unique Roman characters will decrease the activation of Cyrillic alphabet units and the phonemic units activated by them. (For the two versions of this word, the number and identity of shared unambiguous letters is the same.) Activation at the phonemic level will feed to word level units where intralevel inhibitory influences will generate a complex pattern of excitatory and inhibitory influences. Generally, phonemic input from BEHA type words will be enhanced relative to input from VENA type words. (see Figure 1a & b) And, in the terminology of interactive models such as McClelland and Rumelhart (1981), phonologically ambiguous BEHA type words require more operational cycles to settle on a single word unit than do phonologically unambiguous VENA type words (Lukatela, Turvey, & Todorović, 1991).

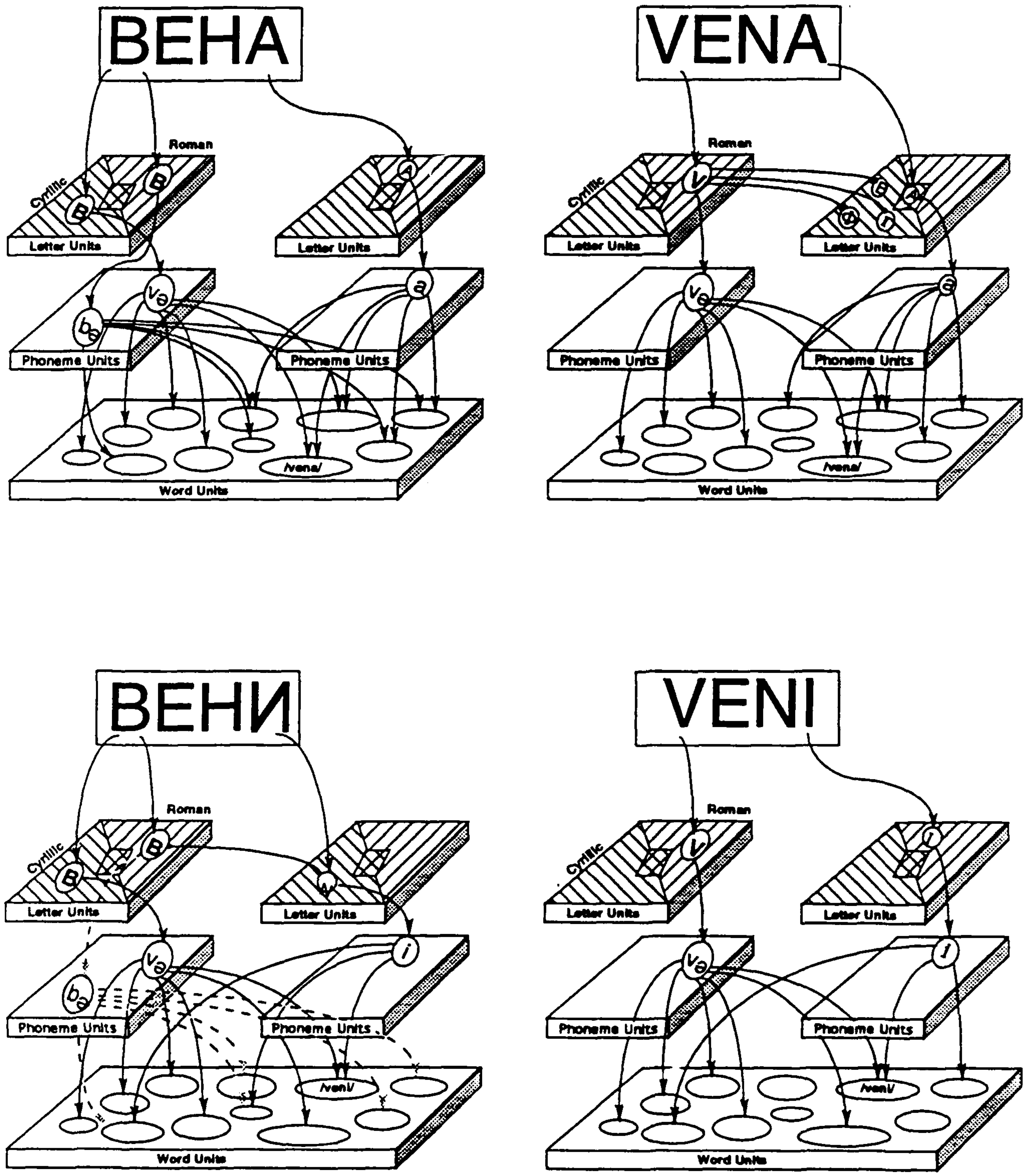

Figure 1. **Patterns of activation for a) BEHA, b)VENA, c) BEHI and d) VENI type words.**

Strings that include a unique letter will, at the letter level, activate the alphabet of the unique letter and (partially) inhibit the letters of the complementary alphabet. Consequently, the composition of strings with ambiguous letters becomes less salient when the string includes a unique letter. For example, the Cyrillic word BEHI which is the dative case of the word meaning "vein" includes a unique letter as its affix in the word final position as well as ambiguous letters in the first and third positions (the first three letters comprise the base morpheme). When activated, Cyrillic I will reduce the potential activation of Roman letters in other positions, that is, the Roman reading of B and H (see Table 1c).

In comparison with BEHA type words where both the Roman and the Cyrillic phoneme units are activated in both the B and H letter positions, the presence of I in BEHI type words will tend to excite the Cyrillic phoneme units of B and H and reduce activation of the analogous Roman units. Consequently, as activation spreads from the phoneme to the word level, the number of highly activated word units will be fewer for BEHI type words than for BEHA type words. Accordingly, lexical decision latencies should be faster for BEHI type words than for BEHA type words and, in fact, latencies for BEHI words were not significantly different than those of VENI type words (see Table 1d).in a lexical decision task (Feldman et al., 1983; Feldman, 1991). Similar effects were also observed in a naming task (Feldman, 1991).

Table 5. *Mean decision and naming latencies (ms) and errors for ambiguous and unambiguous base morphemes with ambiguous and unambiguous affixes (from Feldman, 1991).*

	Base Morpheme		
Affix	Ambiguous	Unambiguous	Difference
lexical decision			
ambiguous	729	671	58
	28	14.3	13.7
unambiguous	677	664	13
	9.2	6.6	2.6
naming			
ambiguous	616	588	28
	25.9	12.5	13.4
unambiguous	626	613	13
	17.6	11.8	5.8

Stated generally, the presence within an isolated letter string of a single character that unequivocally specifies alphabet can bias the activation from letter to phonemic units. This outcome is significant in consideration of the three previous studies where alphabetic manipulations exerted no influence on the processes of word recognition. In the present study, it is evident that an effect of alphabet ambiguity reveals itself when a letter string contains no unique letters to guide alphabet identification. That is, alphabetically-defined visual effects are linked to the phonological characteristics of a word and reveal themselves when a word is phonologically complex. In the last two studies, the domain of alphabet bias is investigated by manipulating the temporal relation between a target and a context that includes unique letters. Transient effects of alphabetically-specified contexts on targets that are and are not comprised exclusively of letters that are shared by both alphabets are examined.

Study 5: Alphabetic manipulations on phonological ambiguity

A fifth line of investigation into the effects of alphabetic bivalence on word recognition and hence a potential source of evidence for an alphabetically-specified orthographic contribution to word recognition entailed primed lexical decision and naming tasks. For target words consisting of phonologically ambiguous strings, plausible related contexts include what the words mean (viz., a semantic associate) and which alphabet yields a word interpretation (viz., alphabetically consistent) as well as a combination of the two.

As described above, some words in Serbo-Croatian can be phonologically ambiguous in either their Cyrillic or their Roman form. For example, BETAP and PAJAC are both phonologically ambiguous because they are composed exclusively of letters that appear in both the Cyrillic and the Roman alphabet. BETAP is a word by its Cyrillic reading (viz., /vetar/ which means "wind") and is meaningless by its Roman reading (viz., /betap/). Conversely, PAJAC is a word by its Roman reading (viz., /pajats/ which means "clown") and is meaningless by its Cyrillic reading (viz., /rajas/). More typically, however, words contain at least one letter that is unique to one alphabet or the other so that a transcription is well-specified with respect to alphabet and phonology (for example, the bold letters of **VETAR** pronounced /vetar/ and **PAJAC** pronounced /pajats/ are unique to their respective alphabets).

In the one experiment (Experiment 2; Lukatela , Feldman, Turvey, Carello & Katz, 1989), targets (either ambiguous or unambiguous) were preceded by a prime that was either alphabetically consistent with the word reading of the ambiguous word or was alphabetically inconsistent with the word reading of that target. Primes were presented for 700 ms with an ISI of 100 ms before the target appeared for 1400 ms. All primes were semantically associated to the critical word targets. That is, BETAP (which means "wind" by its Cyrillic reading) was preceded either by the word for "storm," written in Cyrillic characters (alphabetically consistent) or by the same word written in Roman characters (alphabetically inconsistent) and PAJAC (which means "clown" by its Roman reading) was preceded by the word for "circus," written in Roman characters or by the same word written in Cyrillic characters. Similarly, VETAR (which means "wind" by its Roman reading and cannot be read as Cyrillic) was preceded by the word for "storm," written either in Roman characters or in Cyrillic characters and PAJAC (which means "clown" by its Cyrillic reading and cannot be read as Roman) was preceded by the word for "circus," written either in Cyrillic characters or in Roman characters.

Min F analyses conducted on word latencies between 1500 ms and 400 ms. revealed significant effects of (consistent/ inconsistent) alphabet context and of ambiguity as well as a significant interaction between the two. Alphabet inconsistency of prime and target slowed lexical decision to phonologically ambiguous transcriptions of words by 63 ms and hurt accuracy by 15.9% relative to the consistent condition. That is, phonologically ambiguous BETAP following "storm" printed in Cyrillic characters was faster and more accurate than BETAP following "storm" printed in Roman characters. For phonologically unique transcriptions of those same words, however, alphabet consistency had a nonsignificant effect of 12 ms on latency and 0.2% on accuracy. For example, VETAR following "storm" printed in Roman characters was not significantly faster or more accurate than VETAR following "storm" printed in Cyrillic characters.

The significance of this outcome with respect to understanding the effect of alphabetic context on word recognition is the observation that latency (and errors) for phonologically ambiguous words is dramatically affected by consistency of alphabetic context whereas no analogous effect of alphabet consistency was observed for phonologically unambiguous words. A similar outcome was observed in a naming task (Experiment 4; Lukatela , Feldman, Turvey, Carello & Katz, 1989) where alphabet consistency of prime with the word reading of the target reduced latencies for ambiguous target words by 52 ms and improved accuracy by 4.6% but, for unambiguous words, alphabet consistency had a nonsignificant effect of 8 ms on latencies and 1.6% on errors.

It is important to note that the specification of alphabet by a prior occurring context affects lexical decision and naming of phonologically ambiguous words not only when related word units appear as primes but also when unrelated words and nonwords appear. In fact, the reduction in recognition latencies to ambiguous words in alphabetically consistent contexts relative to alphabetically inconsistent contexts was 86 ms when contexts were defined by unrelated words and was 97 ms when context was defined by a meaningless string of predominantly unique consonants (Experiment 1; Lukatela, Turvey, Feldman, Carello & Katz, 1989). As generally described, a context can bias but will not necessarily restrict processing to one alphabet. That is, all phonemic interpretations permitted by an orthographic string will be activated, at least partially. Finally, because words, both related and unrelated, as well as unpronounceable letter strings can serve as contexts, the effect of context on the activation of letter and associated phonemic units is unlikely to occur at the word level and more plausibly occurs at the linkage between letter and phonemic units.

It is interesting to note that lexical effects can sometimes override the consistent biasing toward one alphabet over another. For example, the word meaning "harem" can be written as either XAPEM which is a Cyrillic form or as HAREM which is a Roman form but the combination HAPEM is meaningless. If the Roman and Cyrillic interpretations are assigned independently for each of the two ambiguous graphemes, this meaningless string can be pronounced in four different ways. One combination is of particular interest: By treating the H grapheme as Roman and the P grapheme as Cyrillic, the word meaning "harem" can be produced from HAPEM. This response constitutes a virtual word. In a lexical decision task, error rates for pseudowords with this structure (i.e., virtual word responses) averaged 42% when they were presented in the context of an unassociated word and increased significantly to 60% in the context of a word that was associated (e.g., the word for "sultan") to the

mixed alphabet reading of this string. And, in the context of an associated prime, correct rejection latencies were slowed by 23 ms relative to the unassociated context (Experiment 4; Lukatela, Turvey, Feldman, Carello & Katz, 1989). In the naming task, 43% of responses to these strings were interpreted as words in the unassociated context and that percentage increased to 63% in the associated context. Similarly, latencies for virtual words named as words were 32 ms faster in the associated context than in the unassociated context (Experiment 5; Lukatela, Turvey, Feldman, Carello & Katz, 1989). Evidently, influences of word level activation on activation at the phonemic level can offset the inhibition of letters belonging to the alphabet not specified by context. That is, in both the lexical decision and the naming task, word level processes can contribute to the pattern of activation in that under some circumstances skilled readers will activate both alphabets in order to interpret a pseudoword as a word.

Alphabet contexts that are consistent across prime and target facilitate recognition of ambiguous target words and sometimes they have a numerically small and statistically nonsignificant effect on unambiguous target words (Experiment 3; Lukatela, Turvey, Feldman, Carello & Katz, 1989). The proposed interpretation of this finding is that the effect of context is to help disambiguate the mapping between letter and phoneme levels. An alternative interpretation is that context could serve to facilitate some later postlexical process. Accordingly, as processing of the context becomes progressively less complete, either in terms of the number of levels stimulated or in terms of the number of elements processed at one level, then strategic and postlexical processing suffers most. By this reasoning, if alphabet biasing is automatic and prelexical, then effects should not vary under experimental conditions that encourage incomplete as contrasted with relatively complete processing of alphabetic information. Alternatively, if alphabet biasing is subject to postlexical strategies and checks then the effect of alphabet may not be evident under conditions that render the context less available.

Study 6: Manipulations of alphabetic accessibility

In principle, alphabetic contexts could exert their influence either early or late in the recognition process. A final methodology for examining the locus of influence of alphabetic context entailed visual presentation conditions in which the availability for processing of alphabetic context was varied by following it with a mask (Lukatela et al., 1991). As in the studies described above, subjects were required to name phonologically ambiguous target words in either Roman or Cyrillic alphabet prime contexts. In one experiment, contexts consisted of 3-5 unique letters which were presented for 70 ms and were followed after an ISI of 30 ms by a target. In this nonmasked condition, results replicated the typical effect of alphabet consistency on naming whereby subjects were 131 ms faster (and 32% more accurate) when both the context and the prime were in the same alphabet than when they were in different alphabets (Experiment 1, Lukatela et al., 1991). Similar results were obtained both when the context duration was reduced to 18 ms and was preceded at an ISI of 0 ms by a masking pattern (Experiment 2) and when the context consisted of a single unique letter (Experiment 4). Evidently, it is not the lexical property of the prime that governs its ability to influence the activation of graphemic and phonemic units.

Effects of alphabetic context have also been observed when the context follows the ambiguous target, and is itself masked so that identifying the alphabetic context and working from there to the target is highly implausible. It is claimed that if processing of the target is disrupted differentially according to linguistic properties of the masked context (and figural properties are held constant), then properties of the masked context must contribute to lexical access for the target and cannot simply influence postlexical processes. In one experiment (Experiments 5; Lukatela et al., 1991), targets consisted of phonologically ambiguous letter strings and their unambiguous alphabet controls and contexts consisted of strings of unique consonants (some of which were repeated) printed in the alphabet that was either consistent or inconsistent with the word reading of the ambiguous letter string. Phonologically, all alphabetically consistent and inconsistent contexts were equivalent. Targets appeared for 40 ms and were followed at an ISI of 0 ms by a context letter string. The context was presented for 40 ms and was followed by a series of hash marks that remained until the onset of the next trial. In that experiment, consistent with previous studies, the difference between correct target identification with alphabetically consistent and inconsistent contexts was 6.98 % for ambiguous targets and 1.46 % for unambiguous targets. This interaction was statistically significant and was

interpreted as evidence that alphabet congruity between target and masked context reduces (and alphabet incongruity augments) the disruption to processing caused by the mask. Because backward pattern masks are assumed to interfere with lexical access, these results were interpreted as prelexical in locus. That is, the benefit associated with alphabetically consistent contexts and targets arises at the level of letter as contrasted with word units.

An interesting prediction that follows from the claim that alphabet effects arise as inhibition at the level of letter units is that when target and subsequent pseudoword mask differ with respect to alphabet, the letters of the mask will be activated relatively slowly because they must overcome prior inhibition from the target. As a consequence, the phoneme units of the target will have more time to activate possible word units. Of course, the effect of masked pseudowords will also be affected by the phonemic similarity of target and mask.

Phonology and alphabet of target and masked prime were manipulated in a backward priming paradigm in which a phonologically unambiguous word target (20 ms) was followed by a pseudowords mask (20 ms) and then by a pattern mask (Lukatela & Turvey, 1990b). Effects of phonological similarity were replicated. Moreover, a significant interaction of phonology and alphabet was obtained. As anticipated, for phonologically dissimilar pairs, alphabetically mismatched targets and masks were identified significantly more accurately than matched pairs. For phonologically similar pairs, there was a nonsignificant trend in the opposite direction. The effect of phonological properties of the mask on target identification suggests enhanced activation of phonemic units activated while processing the target. The interaction suggests a transient inhibition of letter units due to alphabetic status of the mask. That is, under very restricted viewing conditions, alphabetic context can influence the identification of unambiguous letter strings in a manner not unlike its influence on ambiguous letter strings.

CONCLUSION

In six word recognition studies using variations of the lexical decision and naming tasks evidence for alphabetically-defined visual effects was examined. The experimental manipulation common to all studies exploited the bi-alphabetic fluency of skilled readers of Serbo-Croatian and entailed a comparison of presenting context and target strings (or successive presentations of word or pseudoword letter strings) in either the same or in different alphabets. It was observed that relative to alternating alphabet, the preservation of alphabet over successive presentations of a word had no significant effect on recognition. Effects of alphabetic context were evident for target strings that were phonologically ambiguous and were typically slow in both the lexical decision and naming tasks, however. The presence of a letter unique to one alphabet, either in the target string itself or in a prior or later-occurring context was sufficient to diminish and sometimes to eliminate any significant effect of phonological ambiguity. This series of results was interpreted as evidence of mandatory phonological processing in Serbo-Croatian word recognition and suggested a processing architecture efficient at handling two sets of mappings between letter and phoneme.

In the experimental literature, phonological effects are sometimes interpreted as postlexical effects and sometimes interpreted as occurring prior to lexical access. Phonological effects in Serbo-Croatian have been interpreted as reflecting early processes for several reasons including the findings that they occur for both real word and orthographically legal but meaningless pseudoword targets and that the alphabetic context need not be fully processed in order to influence processing of the target. That is, unmasked as well as masked alphabetic contexts have similar effects on phonologically ambiguous letter strings and alphabetic contexts can be words, pseudowords or a single letter. For phonologically unambiguous letter strings, effects of alphabetic context are rare. Finally, rates of target identification under alphabetically matched and mismatched conditions with phonologically mismatched masks suggest that the time course for effects of alphabetic context on unambiguous strings exist but may be quite transient.

The proposed model of a reading mechanism for the skilled reader of two alphabets entails letter, phoneme and word units. Effects in lexical decision are constrained primarily by activity at the word level whereas naming is constrained primarily by activity at the phonemic level. Effects of alphabet arise relatively early in the model and tend to be graded in nature. For example, inhibitory connections between alphabets exist at the letter level so that within a word, activation of a letter unique to one alphabet will tend to reduce the level of activity of letter units in the alternative alphabet. Similarly, the influence of a context that specifies alphabet is to bias the

connections between letter and phoneme toward the designated alphabet. In sum, the processing microstructure for word recognition in Serbo-Croatian includes principles whereby inhibitory connections exist between the letter units of the two alphabets and the systematic covariation of letters and phonemes within each alphabet is realized. Evidence for alphabet affects at the word level are not typically observed.

REFERENCES

Bentin, S., & Peled, B-S. (1990). The contribution of task-related factors to the repetition effect at short and long lags. *Memory & Cognition, 18*, 359-366.

Feldman, L. B. (1981). Visual word recognition in Serbo-Croatian is necessarily phonological. *Haskins Laboratories Status Report on Speech Research, SR-66.*

Feldman, L. B. (1989). *Morphological priming in lexical decision and naming*. Paper presented at the Psychonomic Society, Atlanta, GA, November.

Feldman, L. B. The contribution of morphology to word recognition. (1991). *Psychological Research*, 53.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research, 45*, 55-72.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 573-581.

Feldman, L. B., & Turvey, M. T. (1983). Visual word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance., 9*, 288-298.

Forster, K. I., & Davies, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology; Learning, Memory and Cognition, 10*, 680-698.

Frost, R., Feldman, L. B., & Katz, L. (1990). Phonological ambiguity and lexical ambiguity: Effects on visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 80*, 569-580.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: The encoding bias hypothesis reconsidered. *Memory & Cognition, 8*, 115-123.

Lukatela, G.. Carello, C., & Turvey, M. T . (1990). Phonemic priming with words and pseudowords. *European Journal of Cognitive Psychology, 2*, 373-394.

Lukatela, G., Feldman, L. B., Turvey, M. T., Carello, C., and Katz, L. (1989). Context effects in bi-alphabetical word perception. *Journal of Memory & Language, 28*, 214-236.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition, 8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetic lexical decision. *Language & Speech, 21*, 142-165.

Lukatela, G. & Turvey, M. T. (1990a). Phonemic similarity effects and prelexical phonology. *Memory & Cognition, 18*, 128-152.

Lukatela, G. & Turvey, M. T. (1990b). Automatic and pre-lexical computation of phonology in visual word identification. *European Journal of Cognitive Psychology, 2*, 325-343.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetical word perception. *Journal of Memory & Language, 28*, 237-254.

Lukatela, G., Turvey, M. T. & Todorović, D. (1991). Is alphabet biasing in bialphabetic word perception automatic and prelexical? *Journal of Experimental Psychology; Learning, Memory and Cognition, 17*, 653-663.

Martin, R. C., & Jensen, C. R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory & Cognition, 17*, 505-521.

Masson, M. E. J., & Freedman, L. (1990). Fluent identification of words. *Journal of Experimental Psychology; Learning, Memory and Cognition, 16*, 355-373.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88*, 357-407.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition, 2*, 309-321.

Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 147-195). London: Erlbaum.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General, 114*, 435-450.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399-412.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading* (pp. 81-90). Hillsdale: Erlbaum.

FOOTNOTES

*In D. M. Williams, R. S. Kruk, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities*. Ontario Institute for Education Publication (in press).

†Also State University of New York at Albany.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 163-178

Morphological Analysis of Disrupted Morphemes: Evidence from Hebrew

Laurie Beth Feldman† and Shlomo Bentin††

In concatenated languages such as English, the morphemes of a word are linked linearly so that words formed from the same base morpheme also resemble each other along orthographic dimensions. In Hebrew, by contrast, the morphemes of a word can be but are not generally concatenated. Instead, a pattern of vowels is infixed between the consonants of the root morpheme. Consequently, the shared portion of morphologically-related words in Hebrew is not always an orthographic unit. In a series of three experiments using the repetition priming task with visually-presented Hebrew materials, primes that were formed from the same base morpheme and were morphologically-related to a target facilitated target recognition. Moreover, morphologically-related prime and target pairs that contained a disruption to the shared orthographic pattern showed the same pattern of facilitation as did nondisrupted pairs. That is, there was no effect of disrupting, over successive prime and target presentations the sequence of letters that constitutes the base morpheme or root. In addition, facilitation was similar across derivational, inflectional and identical primes. The conclusion of the present study is that morphological effects in word recognition are distinct from effects of shared structure.

The internal structure of a word plays a key role in its recognition. Whereas much work on visual word recognition has focused on phonology, more recent efforts have focused on aspects of morphology. One experimental task that is sensitive to the morphological components of words is repetition priming. Significant facilitation among visually-presented morphologically related words in the repetition priming variant of the lexical decision task is well documented (Stanners, Neiser, Hernon, & Hall 1979). Generally, responses to targets that are formed around the same base morpheme as their (morphologically-related) primes are faster and more accurate than to targets following unrelated primes. Sometimes, the facilitation with morphological relatives as primes is equivalent to the effect of an identical repetition of the target. Sometimes, it is numerically reduced relative to identical repetitions but is still statistically reliable (Fowler, Napps, & Feldman, 1985). Effects of morphological relatedness with visually presented materials in the lexical decision task have been found across a variety of languages including Serbo-Croatian (Feldman & Fowler, 1987), English (Feldman, 1991a; Fowler et al., 1985) and Hebrew (Bentin & Feldman, 1990) as well as American Sign Language (Hanson & Feldman, 1989; see also Emmorey, 1989). At lags larger than zero or if more than a few seconds separate the second presentation from the first, the pattern of facilitation due to morphological relatedness is distinct from the pattern due to semantic association (Bentin & Feldman, 1990; Dannenbring & Briand, 1982; Henderson, Wallis & Knight, 1984; Napps, 1989). At average lags of 10 items, orthographic similarity of morphologically unrelated prime and target (e.g., pairs such as DIET and DIE) produces neither facilitation nor inhibition (Bentin 1989; Feldman & Moskovljević, 1987;

The research reported here was conducted at Hebrew University in Jerusalem and was supported by funds from National Institute Of Child Health and Development Grant HD-01994 to Haskins Laboratories and from the Israel Foundation Trustees to Shlomo Bentin.

Hanson & Wilkenfeld, 1985; Napps & Fowler, 1987). In short, the repetition priming procedure is a viable tool for studying how the morphological relation among words is represented in the lexicon and how that relation distinguishes itself from other types of similarity.

An examination of morphologically complex words across languages reveals two basic linguistic principles by which such words are constructed. In one, discrete morphemic constituents are linked linearly. There is a base morpheme to which other elements are appended so as to form a sequence. This principle defines a concatenative morphology, of the kind characteristic of English and Serbo-Croatian, for example. In languages with a concatenative morphology, suffixes and prefixes are regularly appended to the base morpheme in a manner that preserves its phonological and orthographic structure. According to the other principle, morphemic units are not just appended to a base form, but also modify its internal structure. This principle defines a nonconcatenative morphology of the kind found in Hebrew, for example (McCarthy, 1981).

In the repetition priming studies of morphological processing conducted with visually-presented English and Serbo-Croatian materials described above, primes and targets were typically constructed around the same base morpheme and only differed with respect to affix. As a result, among morphological relatives, the base morpheme remained intact and unchanged. Exceptions consist of studies that explored effects of changed spelling and/or pronunciation among morphologically related pairs (e.g., HEAL and HEALTH or SLEEP and SLEPT) at long lags (Fowler et al., 1985; Stanners et al., 1979; see also Kempley & Morton, 1982), studies that examined spelling and sound changes among morphological relatives at varying short lags and SOA's (Napps & Fowler, 1987) and a study with German materials that examined umlaut changes (Schriefers, Friederici, & Graetz, 1992). Even in those studies, however, the changes introduced to the base morpheme were relatively minor (e.g., consisting of a vowel or a vowel plus consonant change) as compared to the portion that was preserved. The structure of materials in those studies reflects a general principle of construction for languages with a concatenative morphology. That is, when morphemes are concatenated it is almost always the case that the phonological and orthographic structure of the base morpheme will be preserved among regular morphological relatives (but see Kelliher & Henderson, 1990). A morpheme tends to be a sequence of consonants and vowels that forms a syllable (or several) and concatenative word formation processes do not disrupt the coherence of the morpheme. The implication of this is that in concatenated languages such as English, morphological relatives will tend to have sequences of letters in common. As applied to the construction of materials in the typical repetition priming task where morphologically-related pairs are formed by adding a suffix, the initial portion of primes and targets will tend to be identical.

Nonconcatenative formation processes are less likely to preserve the integrity of the base morpheme. The base morpheme in Hebrew is an abstract form which is called the "root" and is comprised of a string of three (or four) consonants. The root is not a complete phonological unit as it includes no vowels. Superimposed on the root is the "word pattern" which consists primarily of vowels. The root together with a word pattern constitute the word. Some word patterns consist exclusively of vowels and typically, the vowels are infixed between the consonants of the root. Other word patterns include a consonant prefix (e.g., M plus vowel) or a suffix (e.g., vowel plus T) as well. Both the word pattern as well as the root are productive and convey morphological and semantic information (Ornan, 1971). For example, the root **SH-M-N** can take many word patterns including -e-e- to form the noun /Semen/ (which means "oil"), and -a-e- to form the adjective /Samen/ (which means "fat"). Similarly, the root **Z-M-R** can take many word patterns including -a-a, -e-e-, and -i-e-. Note that roots such as **Z-M-R** and **SH-M-N** are productive in that they generate several words in the semantic fields related to singing and oil respectively. Similarly, the word patterns are productive and tend to modify the root in systematic ways (Berman, 1978). For example, the -a-a- word pattern tends to denote an agent, the -e-e- pattern an object, and the -i-e- the past tense of an active verb in the third person singular. Thus, in Hebrew, /zamar/ meaning "a singer," /zemer/ meaning "a song," and /zimer/ meaning "he sang" are all morphologically-related because they share the **Z-M-R** root and are all bimorphemic because they include a word pattern as well as a root.

It is useful to point out that when different roots accept the same word pattern, the semantic information carried by that word pattern is not fully consistent. Specifically, although, the word pattern -a-a- often denotes an agent, it is also sometimes used to denote the past tense singular form of active verbs as well as some adjective

forms. Compare, for example, the contribution of the -a-a- pattern to the root **Z-M-R** (/zamar / meaning "singer") with its effect on the root **L-V-N** (/lavan/ meaning "white"). Similarly, the semantic contribution of the root is not consistent in any simple sense over all morphologically-related items.

The principle of building words in Hebrew, in contrast to that of languages such as English and Serbo-Croatian, dictates that the phonological and orthographic similarity of morphologically-related words in Hebrew will be spread over several syllables. Root morphemes consist of a sequence of consonants and the requisite vowels for a particular word pattern are infixed between the consonants. Consequently, the root morpheme constitutes neither an orthographically nor a phonologically coherent whole. Rather than forming continuous units, morphemes tend to be disrupted and distributed over several syllables.

Alternative accounts of morphological effects

Accounts of morphological effects in word recognition often minimize the role of purely linguistic variables such as the morpheme and rely on orthographic and phonological patterning of letter units or on semantic similarity in conjunction with shared orthographic and phonological structure. For example, Seidenberg (1987) suggested that patterns of high and low probability of transition among sequences of letters could account for (syllabic or) morphological patterning because transitional probabilities of letter sequences that straddle a (syllabic or) morphological boundary tend to be low (bigram troughs) relative to probabilities of sequences internal to a unit. In an illusory conjunction paradigm, subjects who tended to misidentify the color of the target letter were more likely to assign the color of another letter from within the same morphological unit than from an adjacent but different unit. Although this result provides support for orthographic (specifically, bigram) structure in a particular task, it does not negate the influence of morphology in word recognition. Recently, in fact, morphological effects have been demonstrated in a lexical decision task where color boundaries within a word were either consistent or inconsistent with morphological boundaries (Rapp, 1992). Moreover, morphological boundary effects were evident both in words with bigrams troughs at the boundary and in words without troughs. Similar effects have also been reported for compound words (Prinzmetal, Hoffman, & Vest, 1991). Whether or not orthographic factors in morphological processing prove to be relevant for languages with concatenative morphologies such as English, it is difficult to see how they could be adapted easily to nonconcatenated languages such as Hebrew because morphemes are not always coherent units. In sum, the tendency to interpret morphological effects as orthographic patterning makes it essential to examine orthographic influences on morphological processing in a language in which the morpheme is not always an orthographic entity.

The emphasis on orthographic patterning is also evident in morphological parsing models in which the affixes of a morphologically complex word are first eliminated and then the remaining portion of the letter string is matched to candidate entries in a lexicon (e.g., Taft & Forster, 1975). Although affix parsing models may be plausible in languages such as English in which the repertoire of morphological affixes is relatively limited, their practicality is severely compromised in languages with differing morphological structures (cf. Henderson, 1989). In Turkish, for example, sequences of morphological affixes may be appended to one root and the form of those affixes may vary due to phonological factors. Moreover, some affixes may be applied more than once. Consequently, a process of suffix stripping with subsequent analysis of the remainder may have to undergo many iterations before the root can be successfully identified. It has been proposed that for Turkish, priority in morphological analysis of a word goes to the root and only then is its sequence of affixes identified. That process starts at the root and proceeds from left to right (Hankamer, 1989). In contrast, morphological parsing in Hebrew poses special problems because the root morpheme constitutes neither a coherent phonological nor orthographic unit and morphological formation is less systematic.

In a study of morphological analysis using repetition priming with Hebrew materials (Bentin & Feldman, 1990), patterns of facilitation for prime-target pairs that were related by semantic association and by a shared (morphological) root were compared. The study exploited the fact that although words that are constructed around the same root are, by definition, morphologically-related, the semantic relation among morphologically-related forms in Hebrew may vary dramatically. All of the morphological relations of prime and target pairs were derivational in nature. As a consequence, the

meaning of a derived form was not always predictable in any simple way from a semantic analysis of its component morphemes (see Aronoff, 1976). Facilitation due to morphological relatedness was evident at lags that averaged ten intervening items. Moreover, facilitation was equivalent for semantically-close (e.g., kitchen-cook) and semantically-distant (e.g., slaughter-cook) prime-target relatives. That is, the magnitude of facilitation to the target meaning "cook" was equivalent following primes meaning "kitchen" and "slaughter." These findings are, in fact, consistent with the claim based on English materials that at long lags semantic overlap between prime and target does not influence the magnitude of repetition priming (Feldman, 1991a). In summary, when all related words were derivational in nature, and an average of ten items intervened between prime and target, facilitation due to morphological relatedness in the repetition priming task was not sensitive to the semantic similarity of prime and target. This outcome suggests that morphological analysis is not based on the semantic overlap of morphological relatives.

Covariants of morphological structure in Hebrew

With respect to orthographic structure of words in Hebrew, it is important to note that most vowels are represented by optional diacritics placed beneath, above or within the preceding consonant although some vowels are represented by letters. Because words are conventionally written without vowel diacritics, morphologically-complex words that share a root morpheme but differ with respect to word pattern will tend to be orthographically but not phonologically indistinguishable. For example, the words /gever/ and /gavar/ are both written גבר (Note that in contrast to English and to phonemic notation, Hebrew is read from right to left) These words are morphologically related and mean "man" and "overcome," respectively. Because the word pattern is composed exclusively of vowels and because the /e/ and /a/ vowels are represented by optional diacritics, these two words have the same orthographic form as conventionally written. Of course, although both words have phonological forms that are created around the G-V-R root, their phonological forms differ because of the infixed vowels. By contrast, when vowels are written and particularly when one of them is represented by a letter, then the orthographic pattern of the root morpheme, like its phonological pattern is no longer a coherent unit. For example, the sequence מִשְׁמָר is read /mɪʃmar/, meaning "guard" whereas the sequence שׁוֹמֵר, is read /ʃɔmer/, which means "guardian". These words are morphologically related as they share the root r-m-v (SH-M-R) They differ, with respect to phonological form, as well as orthographic form, however, because in one case the letters for the /ɔ/ vowel of the word pattern is infixed between the consonants of the root. In the present study, we use patterns of facilitation for morphologically complex words in the repetition priming task to ask whether the morphological processing of disrupted roots as typically occurs in Hebrew is similar to the processing of continuous roots as typically occurs in concatenated languages.

Linguists distinguish between two types of morphologically complex words. Words that share a base morpheme but differ with respect to *inflectional* affixes are generally considered to be forms of the same word (e.g., CALCULATE, CALCULATED). Words that share a base morpheme but differ with respect to *derivational* affixes are generally considered to be different words (e.g., CALCULATE, CALCULATOR, CALCULATION). As a secondary objective in the present study, we use patterns of facilitation to ask whether inflectional and derivational formations are likely to involve distinct types of representations and/or processing.

Experimental evidence for this linguistic difference has been difficult to obtain in English. One possible reason for the failure to find evidence for the linguistic distinction between inflectional and derivational formations is that the similarity of orthographic form cannot be equated in English. Specifically, because inflectional relatives and derivational relatives tend to differ with respect to length of affix (or because the transitional probability from the final letter of the base morpheme to the initial letter of the affix differs for inflectional and derivational affixes), these comparisons are not appropriate.

In Hebrew, by contrast, it is possible to identify pairs of words that are, respectively, inflectionally- and derivationally-related and are equated with respect to orthographic and phonological similarity to that target. By definition, all such words are morphologically related to each other because they are constructed around the same root morpheme. Words in a pair differ with respect to the word pattern but inflectionally-related and derivationally-related word patterns can be matched with respect to presence (and letter length) of prefixes and/or

suffixes. In this way, the structural similarity to a target of inflectional and derivational relatives can be matched so that types of morphological formations can be compared.

To summarize, the primary goal of the present study was to examine the role of orthographic patterning in morphological analysis. Accordingly, using morphologically nonconcatenated Hebrew materials, the orthographic integrity of the base morpheme across morphological relatives was systematically manipulated in the repetition priming task. Sometimes prime and target presentations preserved the same orthographic form of the root morpheme and sometimes they did not. A secondary goal of the present study was to compare facilitation by inflectional and derivational relatives. Primes and targets shared a common root and primes were either inflectionally- or derivationally-related or identical to the target. Lexical decision latency to the target was compared following morphologically-related and identical primes. A series of three experiments was conducted in an attempt to uncover the contribution to word recognition of orthographic similarity over and above that of morphological relatedness.

EXPERIMENT 1

Across languages, a variety of mechanisms for forming words exist, the most common being the addition or affixation of an element to a base morpheme (Matthews, 1974). Affixation includes three processes, defined by the position relative to the base morpheme, where addition occurs. These include prefixation, suffixation, and infixation in positions initial, final and internal to the base morpheme. Prefixation and suffixation entail the linear concatenation of elements, whereas infixation is nonconcatenative insofar as the integrity of the base morpheme is disrupted. As described above, the characteristic morphological process of Semitic languages, such as Hebrew, relies on a skeleton of consonants into which a pattern of vowels is infixed (although a prefix or suffix may also be appended). The morphological system of Semitic languages is distinguished for its productivity, the manner in which semantic modification of the root occurs among complex forms that share a root, and for the nonconcatenativity of morphemes (Berman, 1978).

As noted above, the orthographic integrity of the base morpheme is generally maintained in English and in Serbo-Croatian but not always preserved in written Hebrew. For processes of infixation, morphological changes typically entail appending different word patterns to a root, where the word patterns specify the requisite vowels of a word. When represented by a letter, vowels in the word pattern necessarily disrupt the sequence of consonants that comprise the root. Consider, for example, the words נָפַל and נֶפֶל and compare them with the target word נוֹפֵל. The target is the present tense of the verb "to fall", in the third person singular (pronounced /nɔfɛl/). The first form is inflectionally-related to the target and is pronounced /nafal/; it is the past tense of the same verb in the same person. The second form is derivationally-related to the target and is pronounced /nɛfɛl/ which means "a dropout". By definition, all three forms are morphological related because they share the same root נ-פ-ל (N-F-L). Note, however, that in the target word, the root morpheme is not continuous. It is disrupted by the vowel O /ɔ/, which is part of the word pattern. Contrast this pattern with that for the words עֶבֶד and עָבַד as compared with the target עֲבָדִים: The target is pronounced /avadim/, meaning "slaves". The first word is inflectionally-related to the target, is pronounced /ɛvɛd/ and is the singular form "slave." The second word is derivationally-related to the target, is pronounced /avad/, which is the past tense, third person singular of the verb "to work." Note that in this case, the orthographic root ע-ב-ד remains intact in all related forms. The orthographic similarity (due to preservation of the orthographic pattern for the root) of the morphological relatives depicted in the latter example is characteristic of all regularly-related pairs in English. In the present experiment with Hebrew materials, the pattern of facilitation due to morphological relatedness of prime-target pairs was compared when the orthographic form of the shared root was disrupted over prime and target presentations (e.g., נ-פ-ל) and when it was intact (e.g., ע-ב-ד).

It has already been demonstrated that in Hebrew, facilitation in the repetition priming task is sensitive to derivational relatedness of prime and target (Bentin & Feldman, 1990). If inflectional and derivational formations in Hebrew are similarly represented in the lexicon then it is anticipated that the magnitude of facilitation in the lexical decision repetition priming task will not vary with type of morphological relation, and a comparison of inflectional and derivational primes is included in the present investigation. It is anticipated that if

orthographic similarity of prime to target is independent of morphological relatedness then the pattern of facilitation for roots that are disrupted and roots that are not disrupted will not differ.

Methods

Subjects. Forty-eight first year students from the Department of Psychology at Hebrew University participated in Experiment 1. All were native speakers of Hebrew. All had vision that was normal or corrected-to-normal and had prior experience in reaction-time studies. None had participated in other experiments in the present study.

Stimulus materials. Forty-eight Hebrew word triplets were constructed. Each included three forms: a target word, a word that was inflectionally-related to it and a word that was derivationally-related to it. All members of a triplet were constructed from the same root morpheme but they differed with respect to word pattern. The orthographic and phonemic overlap of morphologically-related words to their targets was systematically manipulated. Targets consisted of twenty-four verbs in present tense, third person singular and twenty-four plural nouns. For verb targets, the inflected forms were past tense formations (third person singular), and the derived forms were nouns in singular case. In the verb set, the roots were orthographically continuous in both inflected and derived forms, but the roots were disrupted in the target by the infixation of a letter vowel. For the noun targets, the inflected forms were the same nouns in singular and the derived forms were verbs in past tense (third person singular) In this set, the roots were orthographically continuous in targets as well as related forms.

Four types of words preceded each target across experimental lists. Words inflectionally- and derivationally-related to the target, an identical repetition of the target and an (orthographically, phonologically and semantically) unrelated word served as primes The orthographic similarity of the derived and inflected primes to their target was matched within each triplet (All word triplets and their English translations are listed in Appendix A). The unrelated words had the same morphological structure (word pattern) as did the related words (for other targets) although they necessarily had different root morphemes.

Ninety-six pseudowords were constructed by combining meaningless three-consonant root morpheme with real word patterns. Root morphemes in nonwords were not repeated over successive trials so as to enhance the orthographic salience of the words.

Four test orders were assembled. Each list was comprised of 96 words and 96 nonwords. All items were presented with their vowels. The 96 words consisted of the 48 targets and their 48 primes. Twelve targets were preceded by identical repetitions, 12 targets were preceded by derivationally-related primes, 12 were preceded by inflectionally-related primes and 12 targets were preceded by morphologically, orthographically and semantically unrelated word primes. The lag between prime and target varied between 7 to 13 items with an average of 10. The serial position of all target words and pseudowords was identical across test orders. The primes were rotated among the four lists, so that within a list each type of prime was equally represented and, across lists, each target was preceded once by each of the four types of primes.

Procedure. Twelve subjects were randomly assigned to each of the stimuli lists. Thus, the four prime types were compared within subjects across all 48 targets and within stimuli across all 48 subjects. Speed and accuracy were equally emphasized in the instructions.

The stimuli were presented approximately 80 cm from the subject, at the center of a Macintosh monochromatic screen. Each item was exposed until the subject responded or for 2000 ms, whichever came first. The interval between onset of successive stimuli was 2500 ms.

The dominant hand was used for word responses and the nondominant hand was used for nonword responses. Latencies were measured from stimulus onset, to the nearest millisecond using a special software algorithm[1] and errors were automatically registered. Following the instructions, a practice list comprised of 24 items (two identity, two inflectional and two derivational prime-target pairs as well as 12 pseudowords) was presented. After a short pause, the experimental list followed in one block. The complete experimental session lasted about 20 minutes.

Results and Discussion

Lexical decision reactions times more extreme than two SD's from the mean for subjects and for items in each condition were excluded from all analyses. Fewer than 2% of all responses were eliminated by these constraints. Mean lexical decision latencies and errors in Experiment 1 are summarized in Table 1.

Table 1. *Mean lexical decision time and percent errors for targets following morphologically-related and unrelated primes words in Experiment 1 (SEm in parentheses).*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
RT	769	701	709	710
	(14)	(14)	(14)	(13)
Errors	0.5	0.5	0.4	0.5
	(0.3)	(0.2)	(0.1)	(0.2)

The statistical reliability of the repetition priming effect was tested in each task by ANOVA with repeated measures across subjects (F1) and across stimuli (F2). In the lexical decision task the effect of prime type was significant $F1$ $(3,141)=20.72$, MSe=2279, $p<.0001$, and $F2$ $(3,141)=18.67$, MSe=2704, $p<.0001$. Tukey-A post hoc comparisons revealed that whereas all prime types significantly facilitated lexical decision relative to the unrelated condition ($p<.01$), the magnitude of the effect did not differ from one type of prime to another. In particular, it was interesting that facilitation with identity primes was not significantly larger than with inflectional or derivational primes.

The factors of Target continuity (disrupted, continuous), and Prime type (unrelated, identity, inflectional, derivational) were examined in an analysis of variance. This analysis revealed that the effect of target continuity was not reliable $F2$ $(1,46)=0.43$, MSe=17429, $p>.50$.[1] The effect of Prime type was significant but, as suggested by the absence of a reliable interaction between Prime type and continuity $F2$ $(3,138)=0.44$, MSe=2737, $p>.50$, facilitation from morphologically-related primes to orthographically disrupted target did not differ from facilitation to orthographically continuous targets.[2]

Table 2. *Mean lexical decision latency in milliseconds (and SEm) for target words with disrupted and continuous roots following primes in the four priming conditions of Experiment 1.*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
Disrupted	759	699	704	709
	(19)	(14)	(14)	(16)
Continuous	785	704	715	717
	(21)	(13)	(17)	(15)

The error rate on words was very low and did not differ as a function of prime type $F1$ $(3,141)=0.24$, MSe=0.4, $p<.80$. Due to the design of the experiment, facilitation due to repetition of pseudowords could not be analyzed.

Experiment 1 had three important outcomes: a) The magnitude of the facilitation in lexical decision was similar for prime-target pairs whose structure preserved the orthographic continuity of the root and for those where the continuity of the root was disrupted by infixing an additional letter. It was the case that all the disrupted roots were embedded in verb targets whereas all the continuous roots were in nouns and that the derivationally-related primes (but not the inflectionally-related primes) always introduced a change in word class between prime and target. Nevertheless, statistically nonsignificant and numerically small differences between facilitation by inflectional and by derivational relatives were obtained. This outcome suggests that the morphological repetition effect is sensitive neither to similarity of orthographic form between the prime and the target nor to the similarity of word class. b) Significant facilitation for inflectionally- and derivationally-related as well as for identity primes was observed and provided further evidence for morphological analysis in Hebrew. However, the magnitude of the facilitation in lexical decision was not significantly greater for prime-target pairs related by inflection than for pairs related by derivation. Thus, facilitation by repetition priming was not sensitive to the type of morphological relation. c) Finally, facilitation due to morphological relatedness in Hebrew cannot be attributed to repetition of an initial syllable. Although the initial consonant was always unchanged in prime and target, the following vowel did vary. Initial consonant and vowel overlap of prime and target was greater for inflections than for derivations for the nondisrupted targets whereas the vowel never overlapped for the disrupted targets. Nevertheless, the pattern was similar for both.

In conclusion, the results of Experiment 1 replicate effects of morphological relatedness in the repetition priming task when the orthographic integrity of the base morpheme is preserved over prime and target and extends the outcome to cases where the continuity of the root morpheme is disrupted. In addition, it shows that the tendency for enhanced semantic overlap of inflectionally-related prime-target pairs relative to derivationally-related pairs contributes nothing to the pattern of facilitation. Collectively, these results provide no behavioral evidence for a linguis-

tic distinction between morphological types. Moreover, it suggests that effects due to morphological relatedness are not easily interpreted as a composite of orthographic and semantic similarity.

EXPERIMENT 2

The results of the previous experiment revealed morphological analysis in the lexical decision task. Evidently, subjects were sensitive to repetitions of a sequence of consonants that comprises a root morpheme whether or not they form an orthographic unit. Importantly, inflectional relationships and derivational relationships produced the same pattern of facilitation. We assume that this outcome can be interpreted as a failure to find evidence for a psychological distinction between morphological types in Hebrew. Aspects of stimulus construction in Experiment 1 permit an alternative account, however.

In Experiment 1, all nonwords were constructed from a meaningless string of consonants combined with a real word pattern and all words (necessarily) consisted of a meaningful root combined with an appropriate word pattern. Therefore, in order to perform the lexical decision task successfully, it was not logically necessary for subjects to attend to the whole word: an analysis of the *root* would have been sufficient. Consequently, it is possible that the failure to observe a difference between morphologically-related primes with inflectional and derivational word patterns reflected the tendency of subjects to ignore perceptually nonsalient vowel information in this experimental setting. It was essential to show that subjects were, in fact, sensitive to the word patterns that create the distinction between inflectional and derivational formations and this was the intent of the second experiment.

In Experiment 2, the informativeness of word pattern information was enhanced by constructing pseudowords along a different principle. Here, pseudowords consisted of a real root and a real word pattern in an illegal combination. The words consisted of the same items as in the previous experiment. The differentiation between word and pseudowords therefore required the subject to process the word pattern as well as the root. As in the previous experiment, word targets were preceded by identity, unrelated, inflectionally- and derivationally-related primes.

Method

Subjects. Forty-eight first year students from the Department of Psychology at Hebrew University participated in Experiment 2. As in the previous experiment, all were native speakers of Hebrew. All had vision that was normal or corrected-to-normal and all had prior experience in reaction-time studies although none participated in other experiments in the present study.

Stimulus materials. The words used in the present experiment were identical to those used in Experiment 1. There were 48 sets, each comprised of a target, an unrelated word, a derivationally-related word, and an inflectionally-related word. Half of the targets contained orthographically continuous roots and half contained roots that were disrupted by the infixation of the vowel /ɔ/ which is represented by the letter O. Both inflectional and derivational primes always included the full root morpheme in a continuous form, and primes were matched for orthographic similarity with the target.

The 96 pseudowords were constructed using other productive roots that exist in the language. All roots were combined with legal word patterns such that the particular combination of root and word pattern was meaningless. For example, the root ע-ב-ד (A-V-D) was combined with the word pattern -ɔ-a-ut in order to form the phonologically legal but meaningless structure עוֹבְדָנוּת which is pronounced /ɔvdanut/. This manipulation was introduced so as to promote morphological analysis of all letter strings.

The four test orders created for Experiment 1 were modified so that a new set of nonwords was substituted for the old set. In all other respects the materials were identical to those of the previous experiment.

Procedure. Subjects were instructed to make a lexical decision judgment. The procedure as well as the word stimuli were identical those that of Experiment 1 except that the timing software was measured from s hardware device that eliminated the constant that had been added to each latency in the previous experiment.

Results & Discussion

Mean lexical decision latencies were calculated in each condition, across subjects and across stimuli. Errors and extreme reaction times were eliminated according to the constraints described for Experiment 1. Mean reaction times and errors for each conditions are presented in Table 3.

The comparison of the latencies of lexical decisions to target words in the different conditions was based on ANOVA using subjects (F1) and stimuli (F2) as random factors.

Table 3. *Mean lexical decision times in milliseconds and percentage of errors for morphologically-related and unrelated target words in Experiment 2 (SEm in parentheses).*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
RT	628	571	576	578
	(12)	(6)	(6)	(10)
Errors	1.9	1.4	1.9	1.5
	(0.4)	(0.3)	(0)	(0.4)

This analysis showed a significant effect of type of prime [$F1$ (3,141)=9.98, MSe=1826, $p<.0001$ and $F2$ (3,141)=20.71, MSe=1744; $p<.0001$]. Post hoc Tukey-A comparisons of the means indicated that the inflectional, derivational, and identity primes all facilitated lexical decision relative to the unrelated condition, ($p<.01$). As in Experiment 1, the magnitude of facilitation was similar for the three related prime types. The analysis of error scores showed no significant difference due to type of prime $F1$ (3,141)=1.49, MSe=1.54, $p>.14$.

The responses to targets with orthographically disrupted roots and targets with continuous roots were compared by a mixed model ANOVA and are summarized in Table 4. Targets with continuous roots were marginally faster than (different) targets with disrupted roots $F2$ (1,46)=3.13, MSe=8787, $p<.084$. The effect of prime type was reliable and, consistent with the outcome of Experiment 1, there was no interaction between type of prime and target continuity F (3,138)=0.78, MSe=2459, $p>.501$. Because the pattern of facilitation was similar for targets with orthographically disrupted and orthographically continuous roots, these data support the conclusion of Experiment 1 that preservation of orthographic pattern is not a necessary condition for facilitation due to morphological relatedness. Finally, neither for disrupted roots nor for continuous roots were inflectionally-related primes and derivationally-related primes significantly different from each other.

The outcome of the present experiment replicated that of Experiment 1. The magnitude of facilitation in lexical decision was not significantly greater for prime-target pairs related by inflection than for pairs related by derivation. More important, neither was facilitation influenced by the orthographic integrity of the repeated root morpheme. Thus, even when the composition of pseudowords forced subjects to analyze the morphological structure of the items in order to perform the lexical decision task, facilitation in morphological repetition priming was not sensitive to a) type of morphological relation nor to b) preservation (or disruption) of an orthographic pattern for the morpheme across prime and target words.

Table 4. *Mean lexical decision latency in milliseconds (and SEm) for target words following primes with disrupted and continuous roots in the four priming conditions of Experiment 2.*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
TARGET				
Disrupted	640	575	584	595
	(22)	(7)	(9)	(17)
Continuous	602	563	568	561
	(11)	(9)	(7)	(10)

Plausible accounts of facilitation in the repetition priming task have identified response-related (episodic) as well as lexical influences (e.g., Bentin & Feldman, 1990; Bentin & Moscovitch, 1988; Bentin & Peled, 1990; Forster & Davis, 1984; Monsell, 1985). One account of the present results places the locus of facilitation at the level of the root morpheme that is repeated in both inflectional and derivational pairs. Perhaps repetition serves to facilitate the identification of an orthographically and semantically abstract root within the composite root plus word pattern that constitutes a word. Conjointly, facilitation may reflect that was present in our previous experiments. It was the case that lexical decision response to a root was also repeated. That is, roots that were parts of words on their first presentation were parts of words on their second presentations. It never was the case that roots that were parts of pseudowords on their first presentation were parts of words on their second presentation. This redundancy between roots and responses might have facilitated the decision process or the selection between the word and not a word response categories, thereby introducing an additional source of facilitation. In the third and final experiment, the lexical decision associated with a particular root was manipulated over repetitions. The experiment was designed in order to identify an episodic component of facilitation associated with response repetition.

EXPERIMENT 3

Pseudoword structure influences rejection time in the lexical decision task. Caramazza, Laudanna and Romani (1988) reported that Italian pseudowords composed of illegal combinations of real morphemes were harder to reject than pseudowords composed of one legal morpheme and one illegal (nonmorpheme) sequence. Similar results have been reported in English (Katz, Rexer, & Lukatela, 1990). Of course, Italian is a concatenated language like English and morphemes consist of uninterrupted sequences of letters whereas in Hebrew the morpheme root is a more abstract unit. Experiment 3 assesses whether Hebrew pseudowords words formed around a meaningful root pose special problems relative to pseudowords formed around a meaningless string of consonants. Hebrew pseudowords constructed by combining meaningful root morphemes with real word patterns were compared with pseudowords constructed of a meaningless root with a real word pattern.

Experiment 3 also attempts to evaluate response repetition as a source of facilitation in this task. In the experiments reported above as well as in all previously reported repetition priming studies, the lexical status of the prime and the lexical decision to the target were matched so that if the answer to the first was "word" then the answer to the second would also be "word" and if the answer to first was "pseudoword" then the answer to the second would also be "pseudoword." In the present experiment, the effect of morphologically-related pseudoword primes on word targets was investigated. That is, primes and targets were always formed around the same root but, due to illegal combinations of root and word pattern, the lexical status of the prime was not always a real word. Failure to find facilitation when the lexical status of prime and target is not matched would provide evidence for a response-related component to facilitation in the repetition priming task.

The addition of a condition in which pseudoword primes are followed by word targets serves to eliminate another potential problem of interpretation. In the previous two experiments, only words were repeated so that it was possible that subjects used repetition of the root as a criterion for deciding the lexical status of a letter string. That is, if a particular string of consonants had been presented previously then respond "word." By this account, target facilitation following unrelated primes would be over estimated as these were first presentations of that consonant string. Accordingly, targets following pseudoword primes formed from the same root should show facilitation because the root is repeated. By contrast, if targets following unrelated primes (with different roots) and targets following pseudoword primes (repeated roots) do not differ significantly, then it is unlikely that subjects are exploiting repetition of the root per se as a basis for judging the lexical status of a target.

Experiment 3 was designed to differentiate the effect of repeating a root morpheme from the effect of repeating a lexical decision response (cf. Logan, 1989). If facilitation following morphological repetition reflects units for accessing the lexicon rather than lexical processes, then target words that contain a root that was previously presented should be faster than targets whose roots were presented for the first time. Importantly, the lexical status of the word in which the root appeared should have no effect. That is, both word and nonword primes that contain the root morpheme should facilitate targets. On the other hand, if morphological components must activate a lexical entry in order to produce facilitation then roots embedded in pseudowords will not facilitate words with those same roots. Such an outcome could also suggest that relatively late processes of decision and response selection contribute to the pattern of facilitation in the repetition priming task.

Method

Subjects. Forty-eight first year students from the Department of Psychology at Hebrew University participated in Experiment 3. As in the previous experiments, all were native speakers of Hebrew. All had vision that was normal or corrected-to-normal and all had prior experience in reaction-time studies although none participated in other experiments in the present study.

Stimulus materials. The materials from Experiment 1 were modified in the third experiment so that the response for a particular root was not necessarily constant over first and second presentations of that root. The materials for the third experiment were identical to those of the previous two experiments with two exceptions. First, instead of including an identical repetition of each target word, a new prime was constructed. It consisted of an illegal combination of the target root and a word pattern. Accordingly, the correct lexical decision response for these primes was *not a word*. As a consequence of introducing a new

principle for constructing primes, two types of pseudowords occurred within each test order. One type consisted of pseudowords formed by creating an illegal combination of meaningful root and real word pattern. These were pseudoword primes for real word targets and twelve existed in each list. The other consisted of a real word pattern on a meaningless root and these were pseudoword fillers. Both types of pseudowords were presented to each subject so that they could be compared.

Thus, the design of the Experiment 3 was similar to the design of Experiment 1, except that here the identity word primes were replaced by pseudoword primes constructed from the same root morpheme that appeared in the target. Within each of the four test orders, the forty-eight targets were preceded equally often by pseudowords, by inflected primes, by derived primes, and by unrelated word primes. Across test orders, each target was preceded by each type of prime.

Procedure. Subjects were instructed to make a lexical decision judgment and the procedure and instructions were identical to those of the two previous experiments.

Results and Discussion

Mean decision latencies and error rates for Experiment 3 are summarized in Table 5. Errors and extreme reaction times were eliminated according to the same constraints used in previous experiments.

The ANOVA of word latencies revealed a significant effect of type of prime [$F1$ (3, 141) = 5.76, MSe = 2016, $p<.001$; $F2$ (3,141) = 11.21, MSe = 2482, $p<.0001$] although the analysis of error scores did not [$F1$ (3,138)=1.13, MSe=1.14, $p>.33$].

Table 5. *Mean lexical decision times in milliseconds (and SEm) and percentage of errors for morphologically-related, unrelated target words and for pseudowords in Experiment 3.*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
RTs (SEm)	653 (13.5)	647 (10.1)	608 (7.30)	609 (7.60)
Errors	2.4 (0.7)	1.9 (0.5)	1.7 (0.5)	1.6 (0.4)

For latencies, post-hoc Tukey-A revealed that targets preceded by inflectionally- and derivationally-related primes were significantly faster than targets preceded by unrelated words. In replication of previous results, the magnitude of facilitation was similar for inflectional and derivational type primes. Reaction times to targets preceded by pseudoword primes were not significantly different from reaction times to targets preceded by unrelated primes, however. This outcome suggests that when the response to a root was not repeated, repetition of the root *per se* was not sufficient to facilitate (or inhibit) lexical decision. This outcome is important because it suggests that word target responses were not simply facilitated because the same root was repeated during the experimental session. Facilitation necessitated activation of a lexical entry.

Comparison of the meaningful root and meaningless root pseudowords revealed that the presence of a meaningful root delayed rejections of pseudowords by about 200 ms. (698 ms vs. 902 ms, respectively). This difference was statistically significant [$F1$(1,47)=88.5, MSe=11280, $p<.0001$] and, is consistent with the results found in concatenated languages such as Italian and English.

As in the previous experiments, latencies and errors to targets containing disrupted and continuous roots were compared. They are summarized in Table 6. The ANOVA showed that continuous and disrupted target types were not significantly different F 2 (1,46)=1.78, MSe=9960, $p>.18$. In replication of previous experiments, the effect of type of prime was significant but there was no interaction between type of prime and continuity $F2$ (1,138)=0.88, MSe=2640, $p>.44$.

Table 6. *Mean lexical decision latency in milliseconds (and SEm) for target words with disrupted and undisrupted roots in the four priming conditions in Experiment 3.*

	PRIME TYPE			
	Unrelated	Identity	Inflection	Derivation
TARGET				
Disrupted	634 (17)	638 (14)	609 (11)	597 (9)
Continuous	661 (19)	650 (14)	613 (10)	631 (12)

The difference in the lexical decision latency between the two types of pseudowords suggests that during the process of lexical decision, roots

were examined and that readers cannot ignore the meaningfulness of the roots even when they are components of pseudowords. Nevertheless, the presence of a meaningful root in a pseudoword could not facilitate later lexical decision to a word formed from the same root. Because only the pseudoword-word combination was examined, this outcome could suggest that repetition of the episode or particular response is a source of facilitation in the repetition priming task. Alternatively, it is plausible that morphological components must activate a lexical entry in order to produce facilitation at a later point. In any event, it appears that the locus of root facilitation cannot be prelexical.

GENERAL DISCUSSION

In a series of three lexical decision experiments, significant facilitation due to morphological relatedness of prime and target was observed with Hebrew materials. Subjects performed a lexical decision to both prime and target and 7 to 13 items intervened between them. When related primes were matched for overall orthographic similarity to targets, facilitation by inflectional primes was equivalent to facilitation by derivational primes, both of which were statistically equivalent to facilitation by identical repetitions. Similar magnitudes of facilitation for the two types of morphological primes is interesting because forms related by derivation generally tend to be less similar in meaning than forms related by inflection (Aronoff, 1976). Moreover, in our particular experiments, pairs related by inflection were always of the same word-class whereas pairs related by derivation changed word class. Evidently, the facilitation that underlies repetition priming among morphologically-related forms cannot reflect preservation of shared meaning over prime and target. These results are consistent with the claim that at long lags, semantic relatedness *per se* is not a primary source of facilitation in the repetition priming task (Bentin & Feldman, 1990), and support a distinction between facilitation due to associative and morphological relatedness (Henderson, 1985).

Alternative accounts of facilitation between morphologically-related prime-target pairs emphasize the repetition of phonological and orthographic patterns conveyed by a shared morpheme. As described above (see also Berman, 1971), and in contrast to concatenated morphologies such as that of English, morphologically complex words in Hebrew consist of a root morpheme of consonants into which a word pattern is infixed. Consequently, root morphemes are abstract patterns that cannot be realized as unified phonological entities. In the present study, roots were repeated over related prime and target but, because word patterns changed, related words were not associated with a common phonological structure. Nevertheless, facilitation was observed. In conclusion, appreciation of morphological relatedness does not require phonological identity. As applied to the repetition priming task, repetition of a phonological unit is not necessary in order to produce morphological facilitation.

Accounts of morphological effects that emphasize orthographic structure (e.g., Seidenberg, 1987; Seidenberg & McClellend,1989) may be more appropriate for concatenated languages because morphemes tend to be orthographic as well as linguistic units. For example, in English, the base morpheme is typically undisrupted by morphological manipulations.[3] Nevertheless, previous studies in English have demonstrated that for morphologically-related words, the repetition of orthographic form plays only a minimal and statistically insignificant role in the morphological repetition effect (e.g., Napps, 1989; Napps & Fowler, 1987). Similarly in Serbo-Croatian, facilitation in repetition priming was numerically equivalent when prime and target were both written in the same alphabet (e.g., NOGOM - NOGA) and when prime was in one alphabet (e.g., noga) and target was in the other (e.g., NOGA) (Feldman & Moskovljević, 1987; Feldman, in press). In Hebrew, the root is always phonologically and sometimes also orthographically disrupted because of its nonconcatenated structure. The major contribution of the present result is to underscore the limitations of an orthographic account of morphological analysis. This claim is based on the following evidence.

First, the magnitude of target facilitation following morphological relatives was similar to that following identical repetitions although the orthographic similarity of the inflected and derived primes to their matched targets was, by definition, smaller than with identity primes. Second, in all three experiments, the comparison between prime-target pairs with orthographic disruptions to the root and pairs with continuous roots yielded no significant differences. Moreover, in Experiment 3, repetition of the root did not facilitate lexical decision to the target if its first presentation was in the context of a pseudoword, even though the pseudoword was as orthographically similar to the target as were the

related words. These results are consistent with the outcome of a similar study conducted with English materials (Fowler, et al., 1985) in that changes in spelling (and/or pronunciation) had no effect on the pattern of facilitation between morphologically-related prime-target pairs in the repetition priming task. The implication of the above is that facilitation in the repetition priming task in nonconcatenated as well as concatenated languages cannot be attributed to repetition of an overall orthographic form nor to preservation, over successive presentations, of the continuity of an orthographic pattern. In summary, morphological analysis is cannot be tied to orthographic units.

Inflections and derivations are contrasted by linguists as representing two different types of morphological formations. In English, inflectional affixes are few and tend to be composed of three or fewer letters whereas derivational endings can be composed of a more variable number of letters. Moreover, some derivations change the meaning and pronunciation of the base morpheme in a manner that is not characteristic of inflections (Chomsky & Halle, 1968). In Hebrew, it is possible to find inflectional and derivational relatives of a target that modify the structure of the root to a similar degree although they necessarily differ with respect to their semantic similarity to the target. In the present repetition priming study, no differences between inflectional and derivational types of morphological formations were observed. Consistent with the conclusion of Napps (1989) and Napps and Fowler (1987), it is evident that facilitation due to morphological relatedness in the present study does not represent the convergence of semantic, orthographic, and phonological relationships.

Locus of morphological effects

In order to observe morphological facilitation in lexical decision, it is not necessary that orthographic pattern be preserved and this finding has been interpreted to mean that morphological analysis is not tied to an orthographic pattern. Similarly, facilitation patterns are not sensitive to the semantic overlap of prime and target in either this or an earlier study (Feldman, 1992). Because the morphological character of a word cannot be captured by its orthographic and semantic properties, it seems that the morphological structure in general and the Hebrew root morpheme in particular must be represented. A morphological representation in the lexicon has been proposed by several investigators (e.g., Grainger, Cole ; & Segui, 1991).

The claim that morphological effects in word recognition reflect lexical processes is based on several sources of evidence. Typically, effects of repeating a morpheme are numerically larger and statistically more robust for word than for pseudoword prime-target pairs. Significant facilitation for pseudowords in the repetition priming task is unreliable even when the negative lexical decision is repeated over prime and target with the same continuous base morpheme (e.g., Duchek & Neely, 1989; Feldman & Moskovljević, 1987). For example, in the one repetition priming study where Hebrew pseudowords were repeated (Bentin & Feldman, 1990), evidence for facilitation due to repetition with pseudowords depended on the choice of a baseline. Similarly, in at least one study with English materials (Fowler, et al., 1985), evidence of facilitation with pseudowords depended on the number of items intervening between prime and target (see also Scarborough, Cortese, and Scarborough, 1986). For morphologically related word pairs, by contrast, effects tend to be larger in magnitude and manipulations of lag are not significant (Feldman, in press). The results of Experiment 3 also cast doubt on a locus for the morphological facilitation that is independent of the lexicon. If it were possible for subjects to extract a root from both word and pseudowords prior to accessing the lexicon, then the effect on word targets of word and pseudoword primes should have been similar. Analogous effects for word and pseudoword primes were not observed, however.

A second source of evidence that (at least some) morphological effects are lexical in origin is the interaction of morphological with frequency effects. Although it is not the case in repetition priming that (relative) frequency of morphologically-related prime and target had a significant effect (Feldman, 1992), morphological and frequency effects often interact in other recognition tasks. Accordingly, more frequent words are less sensitive to manipulations of morphological structure than are less frequent words. For example, in an experimental production task (Stemberger & MacWhinney, 1986; 1988), the error rate on lower-frequency morphologically-complex forms was significantly higher than on higher-frequency verb forms. Similarly, it has been suggested (Caramazza et al., 1985) that both whole word and morphological units may constitute viable units for accessing the

lexicon but that the availability of the former are constrained by the frequency of the particular surface form.

It is important to point out that the measure of variance included in Table 3 provides no evidence that performance was more variable in the pseudoword prime condition than in the unrelated prime condition. Therefore, an account based on compensatory processes such as facilitation due to repetition of the root being offset by a change of response to that root seems implausible.

Evidence that facilitation due to morphological relatedness is lexical in locus is compelling and fits well with the results of studies that used different experimental paradigms. What is less obvious is how to account for the effect of morphemic composition on pseudoword rejection latencies. Rejection latencies were prolonged for pseudowords that included a meaningful root relative to pseudowords that did not This outcome for real roots in illegal combinations with word patterns could reflect a relatively late and strategic re-evaluation of the decision process analogous to the spelling check necessary for pseudohomophone rejection.

Recently, Grainger et al. (1991) have identified two plausible lexical loci for morphological effects in word recognition. As usually conceived, morphological effects are interpreted as sublexical in origin so that morphological relatedness is represented as a system of facilitatory connections between lexical entries for morphologically-related words or as a pattern of activation among morphological units at a level intermediate between word and letter level units. Whether interpreted as a system of connections between whole word forms or as patterns of activation among shared morphological units, the traditional locus of morphological relatedness is sublexical (but not prelexical) in that it is intermediate between word and letter levels. As noted by Grainger and his colleagues (1991), according to a sublexical account, one might expect to observe inhibition among morphologically-related words because of their shared orthographic structure but this outcome has not been reported. Alternatively, morphological units may be represented at a level above the word so that all words formed from the same base morpheme are linked by facilitatory connections to the morpheme and conversely, from the morpheme back to related words. By the supralexical account, activation spreads from a specific word to its base morpheme and then on to other words that are morphologically-related to it. An extension of the supralexical account is consistent with the claim that facilitation in the repetition priming task with Hebrew materials may reflect the process of extracting the root. from the root plus word pattern combination that constitutes a word (Bentin & Feldman, 1990). It also alleviates the problem of identifying a morpheme which, in Hebrew, is neither a phonological nor an orthographic entity. Segmenting root from word pattern in Hebrew necessarily requires extensive lexical knowledge, therefore the process of root extraction in Hebrew must be distinguished from prelexical processes such as affix stripping (Taft & Forster, 1975). Almost all Hebrew pseudowords have legal orthographic (and phonological) patterns so that their differentiation from words must entail examination of the root and may even include an evaluation of its semantic content. This identification may require extracting the root from the word. It is plausible that when roots are repeated over prime and target words in repetition priming, it is the identification of the root that is facilitated. Of course, even the extraction of a semantically meaningful root from its word context is not sufficient to reliably categorize a string as a word. The combination of root morpheme and word pattern must also be evaluated. It was observed in Experiment 3 that pseudowords composed of a meaningful root in illegal combination with a word pattern were more difficult to reject than pseudowords formed around a meaningless root. Activation from the root could spread down to letter level even in the absence of word level activation and this pattern of activation throughout the system could have the effect of biasing the decision process toward a word response.

In summary, both lexical and postlexical influences may contribute to the pattern of facilitation in the repetition priming task. For lexical decision, response repetition about the lexical status of a particular morpheme in a particular (word or pseudoword) context constitutes a postlexical contribution. Support for the lexical aspect of morphological analysis is tied to the pattern of facilitation in the repetition priming task for word targets. It could arise either sublexically or supralexically. The noncon-catenative morphological structure of Hebrew lends itself to a supralexical representation of morphology. If common morphological units are captured at a level above the word then discontinuities of phonological or orthographic components of a morpheme are no longer problematic. Prolonged latencies for pseudoword

composed of illegal combinations of root and word pattern relative to pseudowords composed from nonroot are also anticipated. In sum, morphological analysis in word recognition is not tied to orthographic form and entails lexical knowledge at either a sublexical or a supralexical level.

REFERENCES

Aronoff, M. (1976). *Word formation in generative grammar.* Cambridge, MA: MIT Press.

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology, 42A,* 693-711.

Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology: General, 117,* 148-160.

Bentin, S., & Peled, B-S., (1990) The contribution of task-related factors to ERP repetition effects at short and long lags. *Memory and Cognition, 18,* 359-367.

Berman, R. A. (1978). *Modern Hebrew Structure.* Tel Aviv: University Publishing Projects.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28,* 287-332.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology, 36,* 435-444.

Duchek, J. M. & Neely, J. H. (1989). A dissociative word-frequency × levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory and Cognition, 17,* 148-162.

Emmorey, K. D. (1985). Auditory morphological priming in the lexicon. *Language and Cognitive Processes, 4,* 73-92.

Feldman, L. B. (in press). Bi-alphabetism and the design of a reading mechanism. In D. M. Willows, R. S. Kruk, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities.* Hillsdale, NJ: Erlbaum.

Feldman, L. B. (1992). Morphological relationships revealed through the repetition priming task. In M. Noonan, P. Downing, & S. Lima (Eds.), *Linguistics and Literacy.* (pp. 239-254). Amsterdam/Philadelphia: John Benjamins Pub. Co.

Feldman, L. B., & Fowler, C. A. (1987). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. *Memory & Cognition, 15,* 1-12.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory & Cognition, 13,* 573-581.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology; Learning, Memory and Cognition, 10,* 680-698.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13,* 241-255.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Memory & Language, 30,* 370-384.

Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process.* (pp. 392-408). Cambridge, MA: MIT Press.

Hanson, V. L., & Feldman, L. B. (1989). Language specificity in lexical organization: Evidence from deaf signers' lexical organization of ASL and English. *Memory & Cognition, 17,* 292-301.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. *Language and Speech, 28,* 269-280.

Henderson, L. (1985). Toward a psychology of morphemes. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 15-72). London: Erlbaum.

Henderson, L. (1989). On the mental representation of morphology and its diagnosis by measures of visual access speed. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 357-391). Cambridge, MA: MIT Press.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 211-224). Hillsdale, NJ: Erlbaum.

Katz, L., Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research, 53,* 25-31.

Kelliher, S., & Henderson, L. (1990). Morphologically based frequency effects in the recognition of irregularly inflected verbs. *British Journal of Psychology, 81,* 527-539.

Kempley, S., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology, 73,* 441-454.

Logan, G. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Matthews, P. H. (1974). *Morphology.* Cambridge: Cambridge University

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12,* 373-346.

Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 147-195). London: Erlbaum.

Napps, S. E. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? *Memory & Cognition, 17,* 729-739.

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research, 16,* 257-272.

Ornan, U. (1971) Binyanim, Ubsisim, Netiyot, and Hataiyot. *HaUniversita, 16,* 15-22. (in Hebrew)

Prinzmetal, W., Hoffman, H., & Vest, K. (1991). Automatic processes in word recognition: An analysis form illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 902-923.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language.*

Scarborough, D. L., Cortese, C. & Scarborough, H. S. (1986). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1-17.

Schriefers, H., Friederici, A., & Graetz, P. (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. Quarterly Journal of Experimental Psychology, 44A, 373-390.

Seidenberg, M. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading.* (pp. 245-263). Hillsdale, NJ: Erlbaum.

Seidenberg, M., & McClelland, J. L. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review, 96,* 523-568.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18,* 399-412.

Stemberger, J. P., & MacWhinney, B., (1986). Frequency and lexical storage of regularly inflected words. *Memory and Cognition*, 14, 17-26.

Stemberger, J. P., & MacWhinney, B., (1988). Are inflected forms stored in the lexicon? M. Hammond & M. Noonan (Eds.) *Theoretical morphology*. San Diego, CA: Academic Press.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638-647.

FOOTNOTES

†Also State University of New York at Albany.

††Department of Psychology, The Hebrew University, Jerusalem.

1We thank Len Katz for developing the software.

2Because the effect of continuity was not significant over items, F values over subjects were not included.

3Exceptions include alternations of strong vowels in pairs such as SING-SUNG and MEET-MET.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 179-190

Phonetic Recoding of Print and Its Effect on the Detection of Concurrent Speech in Amplitude Modulated Noise*

Ram Frost†

When an amplitude-modulated noise generated from a spoken word is presented simultaneously with the word's printed version, the noise sounds more speechlike. This auditory illusion obtained by Frost, Repp, and Katz (1988) suggests that subjects detect correspondences between speech amplitude envelopes and printed stimuli. The present study investigated whether the speech envelope is assembled from the printed word or whether it is lexically addressed. In two experiments subjects were presented with speech-plus-noise and with noise-only trials, and were required to detect the speech in the noise. The auditory stimuli were accompanied with matching or nonmatching Hebrew print, which was unvoweled in Experiment 1 and voweled in Experiment 2. The stimuli of both experiments consisted of high-frequency words, low-frequency words, and nonwords. The results demonstrated that matching print caused a strong bias to detect speech in the noise when the stimuli were either high- or low-frequency words, whereas no bias was found for nonwords. The bias effect for words or nonwords was not affected by spelling to sound regularity- that is, similar effects were obtained in the voweled and the unvoweled conditions. These results suggest that the amplitude envelope of the word is not assembled from the print. Rather, it is addressed directly from the printed word and retrieved from the mental lexicon. Since amplitude envelopes are contingent on detailed phonetic structures, this outcome suggests that representations of words in the mental lexicon are not only phonologic but also phonetic in character.

It is generally assumed that the processing of words in the visual and auditory modalities differs in the initial phase because of different input characteristics, but converges at later stages. Hence, findings regarding the influence of orthographic information on the perception of speech, and findings showing how spoken information visual word perception, may suggest how print and speech are integrated in the mental lexicon.

The present study is concerned with a special form of interaction between the visual and auditory modalities during word recognition. It discusses the possible origins of an illusion of hearing speech in noise caused by simultaneous presentation of printed information.

The convergence of printed and spoken stimuli representations during processing has been previously demonstrated in unimodal studies. It has been shown that lexical decisions to spoken words are facilitated if successive words share the same spelling (Jakimik, Cole, & Rudnicky, 1980). Similarly, Hillinger (1980) has shown that priming effects with printed words were enhanced when primes and targets were phonemically similar. However, the influence of one modality on processing in the other modality can be shown more directly in cross-modal studies. It has been established that printed words can prime lexical decisions to spoken words and vice versa (Hanson, 1981; Kirsner, Milech, & Standen, 1983).

This work was supported in part by a grant awarded to the author by the Basic Research Foundation administered by the Israel Academy of Science and Humanities, and in part by the National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. I am indebted to Liora Yancovitch for her help in conducting the experiments, and to Benny Shanon, Bruno Repp, and three anonymous reviewers for their comments on earlier drafts of this paper.

Similarly, using the naming task, Tanenhaus, Flanigan, and Seidenberg (1980) have demonstrated a visual-auditory interference in a Stroop paradigm. These results were interpreted to show that reading and listening share one lexicon, which allows identical messages to be understood in the two modalities in the same way.

Stronger but more controversial evidence concerning the interaction of the visual and the auditory modalities comes from studies demonstrating cross-modal influence occurring before the completion of input analysis. According to a strongly interactive view, some or all stages of the *perceptual* process in one modality may be influenced by activation in the other modality. For example, it has been suggested that automatic grapheme-to-phoneme activation might occur prior to word recognition, hereby affecting the process of auditory lexical access through sublexical activation in the visual modality (e.g., Frost & Katz, 1989; Dijkstra, Schreuder, & Frauenfelder, 1989). Dijkstra et al. (1989) have shown that a visual letter prime can facilitate the auditory detection of a vowel in a syllable. Similarly, Layer, Pastore, and Rettberg (1990) have reported results showing faster identification of an initial auditory phoneme when congruent visual information was presented simultaneously.

Perceptual cross-modal influences can be shown at levels higher than graphemes and phonemes. In a recent study, Frost et al. (1988) have reported an auditory illusion occurring when printed words and masked spoken words appear simultaneously. Subjects were presented with speech-plus-noise and with noise-only trials and were required to detect the masked speech in a signal detection paradigm. The auditory stimuli were accompanied by print which either matched or did not match the masked speech. Since the noise used in this experiment was amplitude modulated, (i.e., the spoken word was masked by noise with the same amplitude envelope), when a printed word matched the spoken word, it also matched the amplitude envelope of the noise generated from it. Frost et al. (1988) have shown that, whether speech was indeed present in the noise or not, subjects had the illusion of hearing it in the noise when the printed stimuli matched the auditory input. These results demonstrate that subject automatically detected a correspondence between noise amplitude envelopes and printed stimuli when they matched. The detection of this correspondence made the amplitude-modulated noise sound more speechlike, causing a strong response bias. This effect was extremely reliable and appeared for every subject tested. The bias effect did not appear when the printed words and the spoken words from which the amplitude envelopes were generated were merely similar in their syllabic stress pattern, or phonologic structure. These results suggest that the printed words were recoded into a very detailed, speechlike, phonetic representation that matched the auditory information, thereby causing the illusion.

One important finding reported by Frost et al. (1988) relates to the processing of nonwords. When the printed and spoken stimuli were pseudowords (nonwords which were phonotactically regular), the bias to hear speech in the noise in the matching condition was much smaller. This result is of special interest because subjects could not identify the masked spoken stimuli, and therefore were unaware that they consisted of nonwords. Nevertheless, they could not detect a correspondence between a printed letter string and its amplitude envelope if it was not a legal word. One possible interpretation of this outcome is that in contrast to words, the covert pronunciation of nonwords is generated either pre-lexically from the print, or indirectly by accessing similar words in the lexicon. Apparently either process is too slow or too tentative to enable subjects to match the resulting internal phonetic representation to a simultaneous auditory stimulus before that stimulus is fully processed.

However, a more radical interpretation of the words-nonwords differences can be suggested. It is possible that amplitude envelopes are stored as holistic patterns in the lexicon, and are addressed automatically by printed words. According to this interpretation, the bias effect could not have been obtained for nonwords, because nonwords are not represented in the mental lexicon, and their printed forms could not have addressed any stored amplitude envelope. It is important to explore this hypothesis further since it has direct relevance to models concerned with the representations of spoken words in the mental lexicon, and with models of visual lexical access. Models of spoken word recognition often assume that representations of words in the lexicon are phonologic in nature, and that the contact representations generated from the speech wave are abstract linguistic units like phonemes and syllables (See Frauenfelder & Tyler, 1987, for a review). According to the above interpretation,

however, representations of spoken words are maximally rich, consisting not only of abstract linguistic units, but also of detailed phonetic information such as spectral templates, and amplitude envelopes. Amplitude envelopes cannot be considered phonological representations because they do not provide the explicit phonemic or syllabic structure of the word. Rather, they retain some speechlike features and convey mostly prosodic and stress information. A similar nonphonologic approach to the mental lexicon, was advocated by Klatt in his LAFS (Lexical Access From Spectra) model (Klatt, 1979; see Klatt, 1989, for a review; see also Gordon, 1988; Jusczyk, 1985).

This issue is also relevant to current discussions concerning the processing of printed words. Models of visual word perception are in disagreement concerning the extent of phonological recoding during printed word recognition (e.g., Seidenberg, 1985; Van Orden, 1987). One class of models assumes that phonological codes are generated automatically following visual presentation and mediate lexical access (Perfetti, Bell, & Delaney, 1988; Van Orden, Johnston, & Halle, 1988; and see Van Orden, Pennington, & Stone, 1990 for a review). In contrast, it has been suggested that phonological codes are seldom generated during visual word recognition, and that with the exception of very infrequent words, printed words activate orthographic units that are directly related to meaning in semantic memory (e.g. Seidenberg, 1985; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Thus, results demonstrating that a visual presentation of a printed word produces a detailed phonetic representation that includes the word's amplitude envelope, even when the experimental task does not require it, provide support for automatic and rapid phonetic recoding in silent reading.

The aim of the present study was to examine further the hypothesis that amplitude envelopes representations of spoken words are not assembled pre-lexically from the print, but are stored holistically in the mental lexicon, and are addressed directly and automatically by matching printed words following lexical access. The generation of a phonetic representation from the print can theoretically be achieved through a prelexical process that maps representation of graphemes into representation of phonemes by applying grapheme-phoneme correspondence rules, and subsequently by transforming the abstract phonologic structure into a detailed representation for silent or overt reading. This process has been often suggested to characterize the naming of novel words or of nonwords (e.g., Coltheart, 1978). Note that whether the phonologic and phonetic structures are derived by applying grapheme-phoneme correspondence rules (Venezky, 1970), or by analogy (Glushko, 1979) is irrelevant in the present context, since both procedures assume that the phonologic code is generated *prior* to the selection of a lexical candidate (i.e., prior to lexical access). In contrast to this account, the hypothesis forwarded in the present study suggests that possible differences in bias between words and nonwords do not result from the relative speed or ease with which the graphemic structure is transformed pre-lexically into a phonetic code. Rather, they emerge because printed words address a maximally rich lexical representation which contains, among other things, the amplitude envelope of the spoken word. Nonwords, on the other hand, are not represented in the mental lexicon, and therefore cannot address their amplitude envelope.

For this purpose, the present study employed the speech detection task proposed by Frost et al. (1988) and examined whether the bias effect caused by matching print depends on the speed of print processing and on spelling-to-sound regularity, or whether it has a lexical origin. Spelling-to-sound regularity and the speed of generating phonological codes were manipulated by using word frequency and the unique characteristics of the Hebrew orthography.

In the two experiments reported here, subjects were presented auditorily with speech-plus-noise or with noise only trials, simultaneous with a visual presentation of printed Hebrew words. In Hebrew, letters represent mostly consonants, while vowels can optionally be superimposed on the consonants as diacritical marks. Like other Semitic languages, Hebrew is based on word families derived from tri-consonant roots. Therefore, many words share a similar or an identical letter configuration. If the vowel marks are absent, a single printed consonantal string usually represents several different spoken words. Thus, in its unvoweled form, the Hebrew orthography is considered a very deep orthography: it does not convey to the reader the full phonemic structure of the printed word, and the reader is often faced with phonological ambiguity.[1] In contrast, the voweled form is a very shallow writing system. The vowel marks convey the missing phonemic information making

the printed word phonemically unequivocal (but see Frost, in press, for a discussion).

Several studies in Hebrew have established that the presentation of unvoweled print encourages the use of orthographic codes to access the lexicon. In order to assign a correct vowel configuration to the printed consonants to form a valid word, readers of Hebrew have to draw upon their lexical knowledge. The complete phonological structure of the printed word can only be retrieved post-lexically, after one word candidate has been accessed. (Bentin, Bargai, & Katz, 1984; Frost, Katz, & Bentin, 1987). In contrast, the explicit presentation of vowel marks provides the reader with the complete phonemic structure of the word (or nonword). Because the voweled orthography is characterized by grapheme-to-phoneme regularity, the diacritical marks enable the generation of a pre-lexical phonologic code by using simple spelling-to-sound conversion rules. This special characteristic of the Hebrew orthography was exploited in order to investigate whether the bias effect caused by matching print on speech detection in noise is affected by the print's phonologic transparency. Specifically, we examined whether the bias effect is dependent on the presentation or the omission of vowel marks.

EXPERIMENT 1

In Experiment 1 subjects were presented with high- and low-frequency spoken words, as well as with nonwords, which were masked by noise with the same amplitude envelope. In addition, the noises were presented alone. The subjects' task consisted of deciding in each trial whether speech was present in the noise, or whether there was noise only. Simultaneous with the auditory presentation, a printed unvoweled Hebrew letter string appeared on a computer screen. Sometimes the printed word or nonword matched the auditory stimulus, and sometimes it did not. In each of these experimental conditions it was determined whether the print caused a bias to hear speech in the noise.

The purpose of this experiment was three-fold: First, to examine whether the bias effect obtained in the shallower English orthography, can be obtained in the deeper Hebrew orthography. If the effect depends on the speed of generating amplitude envelopes pre-lexically from the print by using spelling-to-sound conversion rules, then the unvoweled Hebrew is at a clear disadvantage. It does not convey explicitly to the reader the full phonemic information necessary for the construction of the amplitude envelope. Although the vowel information can be retrieved from the lexicon following visual lexical access, this process is slower to develop. Indeed, a multilingual comparison of naming latencies (Frost et al., 1987) revealed that naming in unvoweled Hebrew is slower than naming latencies in shallower orthographies like English and Serbo-Croatian. Moreover, in contrast to English and Serbo-Croatian, naming latencies in Hebrew were found to be slower than lexical decisions. This is because the phonemic structure necessary for naming is not conveyed directly by the print, but retrieved from the lexicon (Frost et al., 1987).

Another factor which affects the speed of generating phonetic codes from print is word frequency. Hence, the second aim of Experiment 1 was to examine whether the bias effect, if obtained, depends on word frequency, or merely on word lexicality. If our previous differences in bias between words and nonwords resulted from the speed by which the printed words and nonwords were transformed into a phonetic structure, then one should expect a stronger effect of bias for high-frequency words relative to low-frequency words. This is because it is easier to retrieve the phonetic structure of high-frequency words (as reflected by faster RTs for these words in the naming task). If, on the other hand, the origin of the bias effect is purely lexical, then a bias should be obtained for all words, whether frequent or nonfrequent, but not for nonwords.

Finally, the third aim of the experiment was to examine the bias effect in a mixed design of words and nonwords. Note that in the original study reported by Frost et al. (1988) we employed a blocked design. One serious handicap with our previous blocked design was that subjects knew in advance whether the auditory stimuli were words or nonwords. This might have encouraged the adoption of different strategies for words and for nonwords, hereby causing the differences we obtained in the bias effect. In a mixed design such uniform strategy cannot be adopted. Thus, if our previous results were caused by this methodological factor, then no significant differences in bias between words and nonwords should emerge in the present mixed design, and nonwords would show the effect as well.

Methods

Subjects. Twenty-four undergraduate students, all native speakers of Hebrew, participated in the experiment for course credit or for payment.

Stimulus preparation. The stimuli were generated from 24 disyllabic words and 12 disyllabic nonwords that had a stop consonant as their initial phoneme. The number of phonemes for all stimuli was either four or five. The 24 words consisted of 12 high-frequency words and 12 low-frequency words. Because there are no reliable sources of standard objective word frequency counts in Hebrew, subjective frequencies were assessed by averaging the ratings of 50 subjects on a 1 (least frequent) to 7 (most frequent) scale. The mean ratings of the high- and the low-frequency words were 5.3 and 3.1, respectively. The 24 words were all unambiguous in unvoweled print- that is, their orthographic form represented only one lexical entry. Thus, each letter string could be read as a meaningful word in only one way, by assigning to the consonant one specific vowel configuration. The nonwords were, in fact, pseudowords, that were constructed by altering one or two phonemes of real words. All nonwords conformed to the phonotactic rules of the Hebrew language.

The auditory stimuli were originally spoken by a male native speaker in an acoustically shielded booth and recorded on an Otari MX5050 tape-recorder. The speech was digitized at a 20 kHz sampling rate. From each digitized word, a noise stimulus with the same amplitude envelope was created by randomly reversing the polarity of individual samples with a probability of 0.5 (Schroeder, 1968). This signal-correlated noise retains a certain speechlike quality, even though its spectrum is flat and it cannot be identified as a particular utterance unless the choices are very limited (see Van Tasell, Soli, Kirby, & Widin, 1987). The speech-plus-noise stimuli were created by adding the waveform of each digitized word to that of the matched noise, adjusting their relative intensity to yield a signal-to-noise ratio of -10.7 dB.

Each digitized stimulus was edited using a waveform editor. The stimulus onset was determined visually on an oscilloscope and was verified auditorily through headphones. A mark tone was then inserted at the onset of each stimulus, on a second track that was inaudible to the subjects. The digitized edited stimuli were recorded at three-second intervals on a two-track audiotape, one track containing the spoken words while the other track contained the mark tones. The purpose of the mark tone was to trigger the presentation of the printed stimuli on a Macintosh computer screen.

Design. Each of the high-frequency words, low-frequency words, and nonwords was presented in two auditory forms (1) Speech-plus-noise trials, in which the spoken stimulus was presented masked by noise. (2) Noise-only trials, in which the noise was presented by itself without the speech. Each of these auditory presentations was accompanied by two possible visual presentations: (1) a matching condition (i.e. the same word or nonword that was presented auditorily and/or that was used to generate the amplitude-modulated noise, was presented in print); (2) a nonmatching condition (i.e., a different word or nonword, having the same number of phonemes and a similar phonologic structure as the word or nonword presented auditorily, or that was used to generate the noise, was presented in print). Thus, there were four combinations of visual/auditory presentations for each word or nonword, making a total of 144 trials in the experiment.

Procedure and apparatus. Subjects were seated in front of a Macintosh SE computer screen (9" diagonal, screen size), and listened binaurally over Sennheiser headphones at a comfortable intensity. The subjects sat approximately 70 cm from the screen, so that the stimuli subtended a horizontal visual angle of 4 degrees on the average. A bold Hebrew font, size 24, was used. The task consisted of pressing a "yes" key if speech was detected in the noise, and a "no" key if it was not. The dominant hand was always used for the "yes" responses. Although the task was introduced as purely auditory, the subjects were requested to attend carefully to the screen as well. They were told in the instructions that, when a word or a nonword was presented on the screen, it was sometimes similar to the speech or noise presented auditorily, and sometimes not. However, they were informed about the equal proportions of "yes" and "no" trials in each of the different visual conditions.

The tape containing the auditory stimuli was placed on a two-channel Otari MX5050 tape-recorder. The verbal stimuli were transmitted to the subject's headphones through one channel, and the trigger tones were transmitted through the other channel to an interface that directly connected to the Macintosh, where they triggered the visual presentation.

The experimental session began with 24 practice trials, after which the 144 experimental trials were presented in one block.

Results and Discussion

The indices of bias in the different experimental conditions were computed following the procedure suggested by Luce (1963). Results computed

according to Luce's procedure tend to be very similar to results produced by the standard signal detection computations (e.g., Wood, 1976). However, Luce's indices do not require any assumptions about the shapes of the underlying signal and noise distributions, and are easier to compute relative to the standard measures of signal detection theory. The Luce indices of bias and sensitivity originally named- *lnb* and lnη, but renamed here for convenience *b* and *d* are:

$$b = 1/2 \ln [p(yes/s+n)\, p(yes/n) / p(no/s+n)\, p(no/n)],$$

and

$$d = 1/2 \ln [p(yes/s+n)\, p(no/n) / p(yes/n)\, p(no/s+n)],$$

where *s+n* and *n* stand for speech-plus-noise and noise only, respectively. The indice *b* assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no." For example, according to the above formula, in order to obtain an average *b* of +0.5, the subject must generate on the average 60 percent more positive responses than negative ones. The indice *d* assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance.

The average values for the bias indices in each experimental condition are shown in Table 1 (top). There was a bias to say "yes" in the matching condition for high-frequency and for low-frequency words, whereas there was no bias in the nonmatching condition. The bias effect found for high-frequency words was not stronger than that for low-frequency words. In fact the opposite pattern was obtained. In contrast to the high- and the low-frequency words, there was no bias to say "yes" for nonwords in the matching condition. There was, however, a bias to say "no" in the nonmatching condition.

The bias indices were subjected to a two-way analysis of variance with the factors of word type (high-frequency words, low-frequency words, and nonwords) and visual condition (matching print, nonmatching print). The main effects of word type and visual condition were significant ($F(2,46)=17.3$, MSe = 0.48, $p < 0.001$, and $F(1,23)=22.0$, MSe= 0.64, $p<0.001$, respectively). The two-way interaction was also significant ($F(2,46)=6.5$, MSe=0.19, $p<0.003$). A Tukey post-hoc analysis revealed that the differences in bias between either type of words and between the nonwords were reliable, as well as the difference between the high- and the low-frequency words ($p<0.05$). The apparent greater bias to say "no" in the nonmatching condition relatively to the nonmatching condition for the nonwords was not significant.

Table 1. *Bias indices (b), and (Standard Error of the Means) for high-frequency words, low-frequency words and nonwords, when matching and nonmatching print is presented simultaneously with masked speech. Print is presented unvoweled. The top b indices were averaged for all subjects, whereas the bottom b indices were averaged for the 12 subjects with the highest detectability scores (d).*

	High-Frequency Words	Low-Frequency Words	Nonwords
Match	0.55	0.94	-0.19
	(0.14)	(0.18)	(0.14)
No Match	-0.11	0.02	-0.48
	(0.12)	(0.14)	(0.13)
Average *d* = 0.15 (n=24)			
Match	0.57	0.99	-0.20
	(0.20)	(0.19)	(0.15)
No Match	-0.35	-0.15	-0.67
	(0.17)	(0.13)	(0.17)
Average *d* = 0.32 (n=12)			

The average *d* in the experiment was 0.15. Hence, the signal-to-noise ratio which was employed in the experiment resulted in a very low level of detection.[2] In order to ensure that the obtained pattern of bias was not affected by the low detection level, the subject sample was split in half, and the average *b* indices were recomputed for those subjects with highest *d*. The average *b* for this sample in the different experimental conditions are presented in Table 1 (bottom), and confirm that the bias was unaffected by the level of detection. This outcome is in accordance with results presented by Frost and his colleagues showing significant bias effects over a wide range of signal-to-noise ratios.[3]

The data of Experiment 1 thus reveal that the bias in the visual matching condition was obtained even in the unvoweled Hebrew orthogra-

phy. However, similar to our previous study, this effect can be demonstrated only when legal words are presented in the visual modality. The bias effect was not reduced when the printed words were relatively infrequent. This suggests that the speed by which the phonetic structure of the word is retrieved does not affect the illusion of hearing speech in the noise in the matching condition. [4] The unexpected stronger effect of bias obtained for the low-frequency words may be possibly related to the phonetic features of the words employed. This possibility will be further considered in the General Discussion.

The most significant outcome of the experiment is that there was no bias in the matching condition for nonwords. Since in the present study a mixed design was employed, this effect cannot be attributed to a uniform "set" strategy adopted for the nonwords. Although there was a greater tendency to say "no" when nonwords appeared in the nonmatching condition relative to the matching condition, this tendency was not found to be statistically reliable. These results suggest then, that in contrast to words, the presentation of printed nonwords did not easily invoke a phonetic representation which could be compared to the amplitude envelopes presented auditorily. Hence, the outcome of Experiment 1 lends support to the hypothesis that the bias to say "yes" in the matching condition for words only, regardless of their frequency, results from the automatic retrieval of their amplitude envelopes from the lexicon. This process does not appear to be affected by factors related to the speed of generating a phonetic code.

EXPERIMENT 2

One possible criticism of the results of Experiment 1 is that in the unvoweled Hebrew orthography the phonemic structure of printed words can be retrieved from the mental lexicon following visual access. In contrast, the phonemic structure of nonwords cannot be determined unequivocally, since the printed consonants do not specify how exactly a nonword should be read. It might be argued that this caused the different pattern of bias found for words relatively to nonwords. According to this interpretation, the amplitude envelopes of words were not stored as such in the lexicon, but generated on-line from more abstract phonologic or phonetic structures which were retrieved *post-lexically* for the words. Because nonwords are not represented in the lexicon, and because the complete phonetic structure of the nonwords was not specified by the unvoweled print, no bias was obtained for nonwords.

In order to ascertain that this factor did not affect our previous findings, in Experiment 2 the effect of bias was measured when the printed stimuli were voweled. By adding the diacritical vowels marks to the consonants, the Hebrew orthography is as shallow as other orthographies which have a clear and unequivocal mapping of spelling-to-sound (e.g., Serbo-Croatian). The marks convey the full phonemic information that is necessary to produce a pre-lexical phonologic code for both words and nonwords. Therefore, the explicit presentation of vowels eliminates the superiority of words over nonwords in regard to phonologic and phonetic processing: Phonologic recoding of both words and nonwords can be easily and unequivocally occur through a fast pre-lexical process by applying grapheme-to-phoneme correspondence rules, and a phonetic representation that includes the word's amplitude envelope may be generated subsequently from the pre-lexical phonologic representation. If, indeed, an amplitude envelope can be formed on-line from such pre-lexical representations, then the addition of vowel marks should produce a bias effect for nonwords as well as for words in the matching condition.

Method

Subjects. Twenty-four undergraduate students, all native speakers of Hebrew, participated in the experiment for course credit or for payment. None of the subjects participated in Experiment 1.

The design, procedure, and apparatus were identical to Experiment 1, except that the printed words and nonwords were voweled by adding their diacritical marks.

Results and Discussion

The bias indices are presented in Table 2. As in Experiment 1, there was a bias to say "yes" in the matching condition for high- and for low-frequency words, but no positive bias whatsoever for nonwords. There was no positive bias in the nonmatching condition for words. However, similar to the pattern obtained in Experiment 1, there was a bias to say "no" in the nonmatching condition for nonwords. The *b* indices were subjected to a two-way ANOVA with the factors of word type and visual presentation. The main effects of word type and visual presentation were significant ($F(2,46)=10.8$, $MSe=0.5$, $p<0.001$, $F(1,23)=20.2$,

MSe=0.99, $p<0.001$, respectively). The interaction of word type and visual presentation was significant ($F(2,46)=3.30$, MSe=0.15, $p<0.04$). A Tukey post-hoc analysis revealed that the differences in bias between either type of words and between the nonwords were significant ($p<0.05$). The greater bias for a "no" response found for nonwords, in the nonmatching condition relatively the nonmatching condition, was significant as well. The difference in bias between the high- and the low-frequency words was not statistically reliable.

Table 2. *Bias indices and (Standard Error of the Means) for high-frequency words, low-frequency words and nonwords, when matching and nonmatching print is presented simultaneously with masked speech. Print is presented voweled.*

	High-Frequency	Low-Frequency	Nonwords
Match	0.67	0.84	0.00
	(0.16)	(0.18)	(0.15)
No Match	-0.16	-0.0	-0.50
	(0.15)	(0.15)	(0.13)

Average $d = 0.28$ (n=24)

The results of Experiment 2 suggest that the addition of vowel marks did not produce an effect of bias to say "yes" in the matching condition for nonwords. It could be pointed out that there was a significant greater tendency to say "no" when nonwords appeared in the nonmatching condition relative to the matching condition. However, even if the absolute relative difference between the two visual conditions serves as a measure for the effect, this difference was almost twice as large for words than for nonwords, as revealed by the significant two-way interaction. Thus, although the vowel marks conveyed an unequivocal phonemic structure for the printed nonwords, and allowed the generation of a phonological representation for both words and nonwords, the difference in bias between words and nonwords remained unchanged. This suggests that the phonetic representation that includes the amplitude envelope information was available only for words to influence the subjects' judgment. The overall similarity in the effects of bias in Experiments 1 and 2 is striking. This outcome confirms that the bias is independent of the print spelling-to-sound regularity, and provides additional support for the claim that the effect is lexically mediated.

General Discussion

The present study investigated the source of readers' ability to detect a correspondence between a printed word and its amplitude envelope. Experiment 1 revealed that matching print caused a bias to detect speech in a noise amplitude envelope, even in the unvoweled Hebrew orthography. This effect of bias was demonstrated only for words, whether high- or low-frequency, and not for nonwords. In Experiment 2 we found an identical pattern of bias when the printed words were voweled, and therefore were phonologically unequivocal. All voweled words produced the effect, but not the voweled nonwords. Moreover, the overall difference between the matching and the nonmatching conditions was much larger for words than for nonwords.

The bias to perceive speech embedded in amplitude-modulated noise derives from an automatic detection of correspondence between the printed letter string and the speech envelope related to it. The present study was concerned with how exactly is this correspondence detected. In order to match the visual to the auditory information, subjects had to generate from the print the relevant amplitude envelope. We examined whether this can be done by simply applying spelling-to-sound conversion rules to assemble a phonologic representation, and by generating the envelope on-line from a phonetic structure that is contingent on the phonologic representation derived from the print. The results of both experiments suggest that it cannot. Subjects did not show any bias to detect speech in the noise in the matching condition when nonwords were presented. Note that Frost et al. (1988) did found a small effect of bias for nonword in the matching condition. This small effect for nonwords is reflected in the present study by the greater tendency to say "no" in the nonmatching condition relative to the matching condition. This tendency might be related to the overall lower detectability level obtained in the present study. In any event, the absolute difference in bias in the matching relative to the nonmatching condition was much larger for words than for nonwords.

Although a phonetic representation could have been easily generated from the printed nonwords when they were voweled, the difference in bias

between words and nonwords remained unchanged. This outcome suggests that the bias effect is independent of spelling-to-sound regularity. Moreover, the effect seemed unaffected by the speed of print processing. Since the phonetic representation of low-frequency words is slower to generate from the print, the strong bias effect found for low-frequency words relative to high-frequency words suggests that speed of print processing is not a crucial determinant of the effect. Note that the addition of vowels in Hebrew was previously shown to accelerate the phonologic processing of low-frequency words more than for high-frequency words (Koriat, 1985). Nevertheless, the bias effect found for low-frequency words did not increase in the voweled condition relatively to the unvoweled condition.

The stronger bias obtained for the low-frequency words might be related to the phonetic features of the stimuli employed. The magnitude of the bias effect depends among other things on the distinctiveness (or uniqueness) of the amplitude envelope, that affects the clarity of correspondence between the amplitude envelopes presented auditorily, and the word depicted by the print. It is possible that for some low-frequency words this correspondence was exceptionally clear. Recent results by Frost (submitted) support the conclusion that the bias effect is not affected by word frequency per se. In this study the bias for high- and low-frequency phonological alternatives of heterophonic homographs was examined, with an identical signal-to-noise ratio. The results demonstrated very similar bias effects for the high- and the low-frequency phonological alternatives (0.55 and 0.51, respectively).

Taken together, the results of Experiment 1 and 2 suggest that the effect of bias reported in the present and in previous studies is lexically mediated. We assume that the printed word addressed a lexical entry which contained, among other phonologic and phonetic information, the word's amplitude envelope. Thus, the envelope was retrieved from the mental lexicon. By this view, a strong effect of bias can be shown only if the printed letter string can be related to an existing lexical entry. Nonwords do not satisfy this requirement, and therefore did not produce the effect to the same extent.

The conclusion that envelopes are stored as lexical representations is supported by a recent study that examined the influence of lipreading on detection of speech in noise (Repp, Frost, & Zsiga, 1991). This study examined the effect of a visual presentation of a speaker's face on the detection of words and nonwords in amplitude modulated noise. The results demonstrated that an audio-visual match created a strong bias to respond "yes" when the stimuli were words, whereas no bias emerged when the stimuli were nonwords. In contrast to orthographic information, there is a natural isomorphism between some visible articulatory movements and some acoustic properties of speech. Thus, the relations of articulatory movements to phonological and phonetic structure are nonarbitrary, and the correspondence between articulatory information and amplitude envelopes may be perceived without lexical mediation. Nevertheless, subjects did not produce any bias in the matching condition for nonwords.

The proposal that amplitude envelopes are contained as holistic acoustic patterns in the mental lexicon is consistent with a view that lexical representations of spoken or printed words are not exclusively phonologic. Models of speech perception often assume that the speech processing system transforms the physical acoustic pattern into a more abstract linguistic representation which makes contact with the lexicon during word recognition. Regardless of the nature of this representation (i.e., what specific unit serves for activating a lexical candidate), lexical access is often viewed as a process which mediates access to more abstract linguistic information (e.g., Mehler, 1981; Pisoni & Luce, 1987). Our present results seem to suggest that the information contained in the lexicon is richer. In the present study, the presentation of a printed word resulted in the retrieval of an acoustic template- the word's amplitude envelope- from the lexicon.

At first glance, storing the word's amplitude envelope as a holistic pattern might seem to be without apparent benefits. The envelopes cannot identify a specific lexical candidate. However, they do convey prosodic and segmental information (e.g., speech timing, number of syllables, relative stress, and several major classes of consonant manner), that might help in selecting a lexical candidate among a highly constrained set of response alternatives (Van Tasell et al., 1987). Thus, the amplitude envelope might serve as additional information used by the listener in order to identify spoken words which have several acoustic realizations, or which their phonemic structure was not clearly conveyed (cf. Gordon, 1988). In these cases, a match between the

perceived amplitude envelope and the stored template might confirm the identity of a lexical candidate. Clearly, richer representations do not constitute a parsimonious storage system. Nevertheless, the advantage of a more complex representational system is that it often allows a more efficient performance of the native speaker/listener.

One possible role of amplitude envelopes can be suggested in regard to the psychological distinction between words and nonwords. It is often assumed that positive lexical decisions given to a letter string or to a spoken phonemic sequence are based on their relation to a semantic representation, whereas negative decisions result from the lack of such connections to the semantic network. In other words, positive and negative decisions are related to the meaningfulness of the presented stimuli. The results of the present study suggest possibly a different type of criterion. If words address stored amplitude envelopes and nonwords do not, fast lexical decisions might be based, at least in part, on whether the printed letter string invoked a detailed phonetic representation such as the amplitude envelope. According to this interpretation, one factor that differentiates between words and nonwords, and contributes to the word/nonword differences in the lexical decision task, is the generation of a phonetic code that contains envelope information. This suggestion, however, remains speculative and deserves further investigation.

The present study has additional relevance to old and recent debates concerning the processing of printed words. Models of printed word recognition are in disagreement concerning the extent of phonological recoding during visual word recognition. One important controversy relates to the automaticity of phonologic recoding. It is often assumed that phonologic recoding is very slow to develop, and lexical access occurs (with the possible exception of very infrequent words) directly from the visual structure of the printed words to meaning. This view is supported by results demonstrating that spelling-to-sound regularity affects lexical decisions only for low-frequency words (e.g., Seidenberg et al., 1984). In contrast, several studies have suggested that phonologic information is available very rapidly as part of visual access to the lexicon (Perfetti et al., 1988; Van Orden et al. 1988). Perfetti and his colleagues have shown that the effect of a pseudoword mask on the perception of a target word was reduced if there was a phonemic similarity between mask and target (i.e. "made," "mayd"). Using a different experimental technique, Van Orden (1987); Van Orden et al., (1988) showed that when subjects had to decide whether a visually presented word belonged to a semantic category, they often made errors to homophones or pseudohomophones of category instances (i.e., positive responses were given to a "rows" in the category of flowers, or "sute" in the category of clothing). These results were taken to demonstrate that phonologic recoding occurs automatically and pre-lexically during lexical access.

The results of the present study support the view that phonetic recoding occurs automatically following the presentation of a printed word. What our results teach us is that the processing of a printed word results not only in a pre-lexical phonologic representation but also in a very detailed phonetic speech representation, that is lexical, and includes the word's amplitude envelope. This representation is *automatically* retrieved from the lexicon. Note that in this and previous studies which used the speech detection technique, subjects were not required to respond to the printed information. Nevertheless, they detected automatically the correspondence between the visual stimulus and the speech envelope.

In summary, the present study suggests that the presentation of words in the visual or the auditory modalities results in the generation of a rich array of orthographic, phonologic, and phonetic representations. One of these representation is the word's amplitude envelope. Because each of these representations may contact the mental lexicon, auditory illusions can be caused by visual printed information. The bias to detect speech in noise is caused by matching print because printed information arouses very detailed speech codes.

REFERENCES

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning Memory & Cognition, 10*, 353-368.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Dijkstra, T., Schreuder, R., & Frauenfelder, U.H. (1989). Grapheme context effects on phonemic processing. *Language and Speech, 32*, 89-108.

Frauenfelder, U. H., & Tyler, L. M. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*, 1-20.

Frost, R. (submitted). Phonetic recoding of phonologically ambiguous printed words: Evidence from a speech detection task.

Frost, R. (in press). Orthography and phonology: The psychological reality of orthographic depth. In M. Noonan, P. Downing, & S. Lima (Eds.), *Literacy and linguistics*. Benjamins.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17*, 302-311.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by a simultaneous presentation of print? *Journal of Memory and Language, 27*, 741-755.

Glushko, R. J. (1979). The organization of activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 674-691.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. Perception & Psychophysics, 43, 137-146.

Hanson, V. L. (1981). Processing of written and spoken words: Evidence for common coding. *Memory & Cognition, 9*, 93-100.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: the encoding bias hypothesis reconsidered. *Memory & Cognition, 8*, 115-223.

Jakimik, J, Cole, R. A., & Rudnicky, A. I. (1980). The influence of spelling on speech perception. *Paper presented at the XXI annual meeting of the Psychonomic Society*. St. Louis, Missouri.

Jusczyk, P. W. (1985). Auditory versus phonetic coding of speech signals during infancy. In J. Mehler, M. Garrett, & E. Walker (Eds.). *Perspectives in Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11*, 621-630.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and Production of Fluent Speech*. Hillsdale, NJ: Erlbaum.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. The MIT Press: Cambridge, Massachusetts, London, England.

Koriat, A. (1985). Lexical access for low- and for high-frequency words in Hebrew. *Memory & Cognition, 13*, 37-44.

Luce, R. D. (1963). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Layer, J. K., Pastore, R. E., & Rettberg, E. (1990). The influence of orthographic information on the identification of an auditory speech event. *Paper presented at the 119th annual meeting of the Acoustical Society of America*, Pennsylvania.

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, B295*, 333-352.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25*, 21-52.

Repp, B. H., Frost, R., & Zsiga, E. (1991). Lexical mediation between sight and sound in speech reading. *Haskins Laboratories Status Report on Speech Research, SR-107/108*, 243-254.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 43*, 1735-1736.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S., Waters, G. S., Barnes, M. & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*, 513-520.

Van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound, and reading. *Memory & Cognition, 15*, 181-198.

Van Orden, G. C., Johnston, J. C., & Halle, B.L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

Van Orden, G. C., Pennington, B. F., & Stone, G. 0. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488-522.

Van Tassel, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America, 82*, 1152-1161.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Wood, C.C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60*, 1381-1389.

FOOTNOTES

[*] *Cognition*, in press.

[†] Department of Psychology, The Hebrew University, Jerusalem, Israel.

[1] A demonstration of this form of ambiguity may be portrayed in English by the following example: The consonantal string "bttr" may be read as "better," "butter," "bitter," or "batter," which are meaningful words. In addition, many other vowel configurations could be added to the consonants to form nonwords. The Hebrew reader is faced with this form of phonological ambiguity regularly in the unvoweled orthography. The addition of the diacritical marks specifies uniquely one phonological alternative.

[2] The discriminability indices obtained in the present experiments were lower than those obtained by Frost et al. (1988), with a comparable signal-to-noise ratio. This difference may possibly be attributed to differences in the spoken stimuli from which the envelopes were generated. In the present study the spoken stimuli were recorded by a male speaker, whereas Frost et al. (1988) employed stimuli recorded by a female speaker. The detection of speech in amplitude modulated noise is achieved by perceiving local spectral peaks that rise above the flat spectral level represented by the masking noise. Such peaks are more salient with a female speaker because of the higher frequencies that are characteristic to female voices.

[3] Although Frost et al. (1988) showed significant bias effects over a wide range of signal-to-noise ratios, they found reduced bias indices at the lowest ratios. Hence, it is possible that the bias values obtained in the present study were lower than those obtained by Frost et al. (1988), because of the lower level of detection in the present experiment. Note, however, that the interpretation of the results is unaffected by the overall bias values, since it is concerned with the differences in bias between the matching and the nonmatching conditions.

[4] Throughout this paper the assumption is made that the processing of printed low-frequency words is slower than the processing of printed high-frequency words. This assumption is supported by the well documented frequency effect in visual word recognition, but was not directly examined in the present

study (subjects were not required to convey their decisions with any time constraints). Decision latencies in the speech detection task do not reflect exclusively the speed of processing the printed words, but also the complexity of processing the auditory stimuli. Hence, the monitoring of reaction times in this task does not necessarily portray the speed of processing the printed information. Note, however, that the set of stimuli employed in Experiment 1 and 2 is a subset of the stimuli examined by Frost, Katz, and Bentin (1987) in the lexical decision and the naming tasks. The frequency effect obtained by Frost et al. (1987) was over 100 ms supporting the assumption that the processing of the low-frequency words was indeed slower.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 191-204

Processing Phonological and Semantic Ambiguity: Evidence from Semantic Priming at Different SOAs*

Ram Frost† and Shlomo Bentin†

Disambiguation of heterophonic and homophonic homographs was investigated in Hebrew using semantic priming. Ambiguous primes were followed by unambiguous targets at 100 ms, 250 ms, and 750 ms SOA. Lexical decision for targets related to the dominant phonological alternatives of heterophonic homographs were facilitated at all SOAs. Targets related to subordinate alternatives were facilitated only at SOAs of 250 ms or longer. When the primes were homophonic homographs, semantic relationship facilitated lexical decision to targets at all SOAs regardless of the dominance of the meaning to which the targets were related. These data can be accounted for by assuming multiple lexical entries for heterophonic homographs, single lexical entries for homophonic homographs and phonological mediation of accessing meanings. Language specific factors probably account for the long lasting activation of subordinate meanings.

Several studies of lexical disambiguation suggested that all the meanings of a homograph may be automatically activated. One experimental procedure used to demonstrate access to multiple meanings is semantic priming. It has been reported that homographs embedded in sentences facilitate lexical decisions for related targets even if these targets are related to meanings which are different than those implied by the sentence context (e.g., Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). These results were interpreted as supporting an exhaustive, context-independent model of lexical access for homographs, according to which, all possible meanings of one homograph are retrieved in parallel. An alternative view is that contextual information affects lexical processing of homographs at an early stage, selecting only meanings which are contextually appropriate (e.g., Schvaneveldt, Meyer, & Becker, 1976; Glucksberg, Kreuz, & Rho, 1986).

A third approach combines features of both previous views into an ordered access model. This model posits exhaustive access which does not occur in parallel, but is determined by the relative frequency of the two meanings related to the ambiguous word (e.g., Duffy, Morris, & Rayner, 1988; Forster & Bednall, 1976; Hogaboam & Perfetti, 1975; Neil, Hilliard, & Cooper, 1988; Simpson, 1981; and see Simpson, 1984, for a review). Hogaboam and Perfetti (1975) have demonstrated that whatever the biasing context, the dominant meaning of a homograph is retrieved first. Evidence for an ordered access was also presented by Simpson (1981), who showed that in a nonbiasing context, only targets which were related to the dominant meaning of an ambiguous word, were primed. Similarly, differential activation of high- and low-frequency meanings of ambiguous homographs was also demonstrated with event related potentials (Van Petten & Kutas, 1987), and by monitoring eye movements (Duffy et al., 1988; Rayner & Frazier, 1989).

If more than one meaning of a homograph can be retrieved even if it appears in a biasing sentence context, multiple-meaning access should

This work was supported by a grant awarded to the first author by the Basic Research Foundation administered by the Israel Academy of Science and Humanities, and by the National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. We are indebted to Pazit Mey-Raz and Michal Kampf for their help in conducting the experiments, and to Cyma Van Petten, Greg Simpson, Charles Perfetti, Keith Rayner, and three anonymous reviewers for their criticism on earlier drafts of this paper.

be the rule for homographs presented in isolation. This hypothesis was confirmed by Holley-Wilcox and Blank (1980), who found that polysemous primes (e.g., BANK) facilitated lexical decisions to targets related to all of their meanings. Holley-Wilcox & Blank (1980) interpreted their results as supporting the parallel-access model. More recently however, Simpson and Burgess (1985) reported evidence for an ordered access model for isolated homographs. They have shown that in the case of isolated homographs the most frequently used (dominant) meaning is accessed first, while the less frequently used (subordinate) meaning is accessed relatively later.

Most studies of lexical ambiguity focused on homophonic homographs (i.e., letter strings that have a single pronunciation but two or more meanings, e.g., BANK). However, homophonic homographs are not the only forms of word ambiguity. Ambiguity can exist also in the relationship between the orthographic and the phonologic forms of a word. For example, in contrast to "BANK," the printed letter string "WIND" has two different pronunciations, each of which has a different meaning. In a recent study, Frost, Feldman, and Katz (1990) examined the effect of phonological ambiguity in Serbo-Croatian. Subjects were presented simultaneously with printed and spoken words, and were required to determine whether they matched. Phonological ambiguity was produced using letters which represented different phonemes in the Cyrillic and Roman alphabets. The results showed that matching phonologically ambiguous printed words with their spoken realizations was delayed relative to the matching of unambiguous printed patterns in which only letters unique to one alphabet were used. This delay was significantly larger when the ambiguous print was matched with the less frequent spoken alternatives than when it was matched with the more frequent spoken alternative. Frost et al. (1990) suggested that these results support a multiple access model in which dominant alternatives reach a higher level of activation. The effect of phonological ambiguity was examined in English as well. Carpenter and Daneman (1981) have demonstrated that the duration of eye fixations on heterophonic homographs was longer when the phonological alternative implied by the semantic context was a low-frequency word than when it was a high frequency word. In a direct comparison between heterophonic and homophonic homographs, Kroll and Schweickert (1978) found that heterophonic homographs like "wind" take longer to name than homophonic homographs. These results suggest that in English, as in Serbo-Croatian, heterophonic homographs are processed differently than homophonic homographs. However, both in English and in Serbo-Croatian heterophonic homographs form a small and perhaps non-representative group of words.

The unvoweled Hebrew orthography presents an opportunity to examine the process of disambiguating the meaning of heterophonic homographs. In Hebrew, letters represent mostly consonants while vowels can optionally be superimposed on consonants as diacritical marks. In most printed material, (except for poetry, holy scriptures and children's literature), the vowel-marks are usually omitted. Since different vowels may be added to the same string of consonants to form different words, the Hebrew unvoweled print cannot specify a unique phonological unit. Therefore, a printed letter string is very frequently phonologically ambiguous, representing more than one word, each with a different meaning.

In a previous study (Bentin & Frost, 1987) we examined the influence of semantic and phonologic ambiguity on lexical decision and on naming isolated Hebrew words. We found that lexical decisions for unvoweled ambiguous consonant strings were faster than for any of the high- or low-frequency voweled (therefore disambiguated) meanings of the same strings. In contrast, naming ambiguous unvoweled words was as fast as naming the high-frequency voweled alternative, whereas naming the low-frequency alternative was significantly slower. On the basis of these and previous results (Bentin, Bargai, & Katz, 1984), we suggested that lexical decisions for unvoweled Hebrew words are generated prior to the process of phonological disambiguation, probably on the basis of orthographic familiarity (cf. Balota & Chumbley, 1984; Chumbley & Balota, 1984, Seidenberg, 1985). This suggestion also accommodates previous data demonstrating that orthographic information is used for lexical decisions and naming more extensively in Hebrew than in other languages (Frost, Katz, & Bentin, 1987).

In contrast to lexical decision, naming necessarily requires the selection of one phonological alternative of the ambiguous letter string. The significant delay in naming the low-frequency voweled alternative relative to the unvoweled and the high-frequency forms of the same letter string, led us to support the ordered-access model for the retrieval of phonological information. Consequently, we suggested that, when confronted with phonologically ambiguous letter strings, readers retrieve

the high-frequency phonol- -ical structure first. The naming task, however, cannot disclose covert phonological selection processes. In particular, na- -ing does not reveal whether phonological alternatives, other than the reader's final choice, had been accessed during the process of disambiguation. For example, in our previous study, subjects overtly expressed only one phonological structure, more often the high-frequency alternative. However, we could not determine whether alternative words were generated but discarded during the output process, or whether only one word was generated from the print. Moreover, although each phonological form was related to a different meaning, naming does not necessarily imply access to semantic information. Therefore, although our previous results supported a frequency-ordered retrieval of phonological alternatives, a more direct measure was necessary to examine whether more than one meaning of a het- -ophonic homograph is automatically acti- -d, and whether this access is ordered by t- -ative frequency of each meaning.

In the present paper we addressed this question using a semantic priming paradigm similar to that used by Simpson and Burgess (1985). Isolated ambiguous consonant strings were presented as primes and the targets were related to only one of their possible meanings. We assumed that if a specific meaning of the prime is initially accessed, lexical decision for targets that are related to that meaning should be facilitated.

A second question addressed in the present study refers to the time course of activation of dominant and subordinate (i.e., high-and low-frequency) meanings of phonologically ambiguous letter strings. Several studies in English have shown that in a sentence context, the subordinate meaning is active only during a limited period of time (Seidenberg et al. 1982; Van Petten & Kutas, 1987). Similar results were found also for isolated homographs (Kellas, Ferraro, & Simpson, 1988; Simpson & Burgess, 1985). In particular, Simpson and Burgess (1985) found that an SOA of 16 ms between prime and target was sufficient to facilitate lexical decisions for targets related to the dominant meaning, but not for targets related to the subordinate meanings. Relatedness to the subordinate meaning facilitated lexical decisions only when the SOA between prime and target ranged from 100 to 300 ms. The fast decay of the subordinate meaning was explained in that study, by assuming that the limited capacity attention system (Neely, 1977), must focus on only one meaning, and in the absence of disambiguating context, the dominant alternative is usually chosen (see also Kellas et al., 1988). However, since in Hebrew, several phonological units are activated in addition to several semantic nodes, it is possible that the activation of both dominant and subordinate alternatives lasts longer. This might happen, for example, if the retrieval of different phonological units results in more extensive lexical processing. In the present study we examined this possibility by manipulating the SOA between the ambiguous primes and the targets.

EXPERIMENT 1-A

In Experiment 1-a we presented subjects with unvoweled heterophonic homographs as primes. Applying different vowel patterns, each prime could be read both as a high- and as a low-frequency word. In each trial the prime was followed by a word or by a nonword target at 100 ms or 250 ms SOA. Subjects were instructed to read the primes silently and to make lexical decisions to the targets. Across subjects, each target was either unrelated to its prime, or related to the dominant or to the subordinate meaning of the ambiguous prime. Facilitation of lexical decisions in any related condition (relative to the unrelated condition) was considered evidence for accessing the related meaning of the prime.

Method

Subjects. Forty undergraduate students, native Hebrew speakers, participated in the experiment for course credit or for payment.

Stimuli. The primes were 40 ambiguous consonant strings which represented both a high- and a low-frequency word. In the absence of a reliable frequency count in Hebrew, we estimated the subjective frequency of each word using the following procedure: From a pool of 100 ambiguous consonant strings we generated two lists of 100 voweled words each. Each list of disambiguated words contained only one form of the possible realizations of each homograph. Dominant and subordinate meanings were equally distributed between the lists. Both lists were presented to 50 undergraduate students, who rated the frequency of each word on a 7-point scale from very infrequent (1) to very frequent (7). The rated frequencies were averaged across all 50 judges. Each of the 40 homographs that were selected for this study represented two words that differed in their rated frequency by at least 1 point on that scale. The validity of this selection was then tested by naming: Twenty four subjects were presented with the unvoweled homographs, and their vocal responses

were recorded. We measured the relative dominance of each phonological alternative as reflected by the number of times it was actually chosen and pronounced by the subjects. Only those homographs whose frequency judgments coincided with the results obtained in the naming task (i.e., at least 66% of the subjects chose to name the phonological alternative that had a higher frequency rate), were used in the experiment.

Two targets were associated to each selected homograph. One target was semantically related to its dominant meaning, and the other to its subordinate meaning. The targets were all unambiguous (i.e., even without vowel marks they represented only one word). In order to ensure similar semantic relatedness for the dominant and the subordinate meanings, the semantic relation of primes and targets was rated by the same 50 judges on a 7-point scale, from unrelated (1) to highly related (7). The means of those ratings were 5.2 for the dominant meanings, and 5.3 for the subordinate meanings. Each of the 80 targets was also paired with an unrelated prime. The unrelated primes were 40 heterophonic homographs selected from the original pool and different than those used in the "related" conditions. Because none of their possible readings was related to the targets, and because dominance is irrelevant in the unrelated condition, the same prime preceded the targets used in the dominant and the subordinate related conditions. Hence, there were only 40 different ambiguous primes in the unrelated conditions which were rotated across subjects. In addition to the word-word pairs, 80 word-nonwords pairs were introduced as fillers. The words were heterophonic homographs different than those contained in the original pool. The nonwords were consonant strings that have no meaning in Hebrew regardless of vowel configuration. An example of related and unrelated prime-target pairs is presented in Figure 1.

Design. There were eight experimental conditions: Different targets were related to the dominant or to the subordinate meanings of the ambiguous primes; each of the related targets was also presented in an unrelated condition. In each of these four possible pairings, the SOA between primes and targets was either 100 or 250 ms. Four lists of words were formed: Each list contained 10 prime-target pairs in each of the eight experimental conditions and 80 word-nonwords fillers. The prime-target pairs were rotated across lists by a Latin Square design: related pairs in one list, were unrelated in another list, pairs which appeared with a prime/target SOA of 100 ms in one list, appeared with SOA of 250 ms in another list, etc. The purpose of this rotation was to present the targets that were related to the dominant meanings of the primes and the targets that were related to the subordinate meanings of the primes, in both the related and the unrelated conditions, at all SOAs, yet avoiding repetitions within a list. Hence, each target word served as its own control for the measurement of semantic facilitation in an across-subjects design (see Figure 1).

Unvoweled prime		מלח (MLCH)		
Phonological alternatives	Dominant		Subordinate	
	MELACH		MALACH	
semantic meaning	"salt"		"sailor"	
Condition	Related	Unrelated	Related	Unrelated
Prime	MLCH	KLV ("dog")	MLCH	KLV ("dog")
Target	"sugar"	"sugar"	"ship"	"ship"

Figure 1. **Example of related and unrelated prime-target pairs in unvoweled Hebrew.**

Procedure and apparatus. The subjects were tested individually. They were instructed to read the primes and to make lexical decisions only for the targets by pressing a ("word" or a "nonword" response key. The dominant hand was always used for "word" responses. All stimuli were presented at the center of a Macintosh computer screen (bold Hebrew font, size 24). The subjects sat approximately 70 cm from the screen, so that the stimuli subtended a horizontal visual angle of 4 degrees on the average. A trial began with the presentation of the prime which was replaced by the target at the end of the respective SOA period. The target was continuously exposed until a response was recorded. The inter-stimulus interval was 2500 ms from subject's response to the onset of the following prime. Each session started with 16 practice trials. The 160 test trials were presented in one block.

Results

Means and standard deviations of RTs for correct responses were calculated for each subject in each of the eight experimental conditions. Within each subject/condition combination, RTs that were outside a range of 2 SDs from the respective mean were excluded, and the mean was recalculated. Outliers accounted for less then 5% of all responses. This procedure was repeated in all six experiments in the present study.

RTs and errors in the different experimental conditions are presented in Table 1. Lexical decisions to targets related to the dominant meanings of the ambiguous primes were faster than to unrelated targets, at both 100 and 250 ms SOA. In contrast, lexical decisions to targets related to the subordinate meanings were faster than responses to unrelated targets only at 250 ms SOA. At 100 ms SOA, lexical decisions to related targets were apparently slower than lexical decisions to unrelated targets.

The statistical significance of those differences was assessed by an analysis of variance (ANOVA) across subjects ($F1$) and across stimuli ($F2$), with the main factors of semantic relatedness (related, unrelated), dominance of prime-meaning (dominant, subordinate), and SOA (100, 250 ms). The main effects of relatedness, dominance, and SOA were significant: RTs to related targets were faster than to unrelated targets [$F1(1,39)=22.0$, MSe=1789, $p<.001$; $F2(1,39)=15.7$, MSe=2655, $p<.001$]; RTs to targets that referred to the dominant meaning of the prime in the related condition were faster than RTs to targets that referred to the subordinate meaning of the prime [$F1(1,39)=14.6$, MSe=2373, $p<.001$; $F2(1,39)=5.75$, MSe=7509, $p<0.02$]; and RTs at 250 ms SOA were faster than at 100 ms SOA [$F1(1,39)=63.9$, MSe=2315, $p<.001$; $F2(1,39=27.0$, MSe=5319, $p<.001$].[1] Relatedness interacted with dominance [$F1(1,39)=5.62$, MSe=2119, $p<.001$; $F2(1,39)=3.16$, MSe=3594, $p<.08$], and with SOA [$F1(1,39)=14.0$, MSe=1256, $p<.001$; $F2(1,39)=10.5$, MSe=2332, $p<.002$]. The interaction of SOA and dominance was not significant ($F1$, $F2$ <1.0). The three-way interaction was significant in the subject analysis [$F1(1,39)=4.0$, MSe=2191, $p<.05$], but only approached significance in the stimulus analysis [$F2(1,39)=2.7$, MSe=4868, $p<0.10$]. The three-way interaction seems to have resulted in part from greater RTs differences between SOAs for unrelated dominant primes (37 ms) than for unrelated subordinate primes (17 ms). We do not have an explanation for this difference.

Table 1. *Reaction times and (percentage of errors) to related and unrelated targets in the different experimental conditions with phonologically ambiguous (unvoweled) primes (Experiments 1-a and 1-b).*

	Dominant Primes			Subordinate Primes			Nonwords	
SOA	100	250	750	100	250	750	1-a	1-b
Unrelated	715 (9%)	678 (8%)	692 (8%)	718 (12%)	701 (12%)	714 (10%)	754 (11%)	778 (8%)
Related	684 (10%)	639 (7%)	658 (8%)	739 (10%)	669 (13%)	692 (11%)		
Priming Effect	+31	+39	+34	-21	+32	+22		

To elaborate the three-way interaction, and because we were concerned with the different patterns of facilitation for the dominant and the subordinate meanings at the short and the longer SOAs, we conducted separate analyses of the relatedness and dominance effects at each SOA. These respective ANOVAs showed that relatedness interacted with dominance at 100 ms SOA [F 1(1,39)=13.6, MSe=1502, p<.001; F2(1,39)=8.4, MSe=2905, p<.006], but not at 250 ms SOA [F1, F2 (1,39)<1.0]. A Tukey-A post hoc analysis of the interaction at 100 ms SOA revealed that the difference between unrelated targets and targets related to the subordinate meanings of the homographs was not significant, whereas, lexical decisions for targets related to the dominant meaning of the homographs were faster than to unrelated targets.

The differences in error rates between the various experimental conditions did not produce significant effects.

EXPERIMENT 1-B

A more complete description of the time course of activating the dominant and subordinate meanings of heterophonic homographs required examination of the semantic priming effects at an SOA longer than 250 ms. This condition could not be included in the first part of the experiment because the total number of stimuli used in our rotated within-subjects design did not permit an additional division.[2] Therefore, this condition was examined in a second group of 40 subjects sampled from the same population of undergraduates as in Experiment 1-a.

The stimuli, design, and procedure were similar to those used in Experiment 1-a, except that the SOA between primes and targets was 750 ms. To make the structure of the stimulus lists as similar as possible to the previous experiment, we introduced as fillers an identical number of heterophonic homographs with a shorter SOA of 250 ms. Moreover, because subjects encountering only long delays between primes and targets might actively invoke the two phonologic alternatives, whereas subjects encountering both long and short delays might not, a second purpose of the fillers with the shorter SOAs was to prevent subjects from developing this search strategy.

Results

RTs were faster for related targets than for unrelated targets, and for targets related to the dominant meaning of the prime than for targets related to the subordinate meaning (Table 1). The statistical significance was assessed in a two-way analysis of variance across subjects (F1), and across stimuli (F2). The main factors were semantic relatedness (related, unrelated), and dominance of prime-meaning (dominant, subordinate). The ANOVA showed that both main effects were significant: [F1(1,39)=24.3, MSe=1286, p<0.001, a... F2(1,39)=14.6, MSe=1765, p<0.001 for seman... relatedness, and F1(1,39)=10.3, MSe=3123, p<0.002, and F2(1,39)=11.6, MSe=3378, p<0.002, for dominance of the prime-meaning]. The interaction of the two factors was not significant [F1, F2 (1,39)<1.0]. Planned comparisons revealed that RTs to targets related to the subordinate alternatives of the prime-meanings were significantly faster than in the unrelated condition [t(1,39)=2.54, p<0.01]. The pattern of semantic facilitation obtained for the fillers with 250 ms SOA was similar to the pattern obtained with the identical SOA in Experiment 1-a (33 ms facilitation for targets related to the dominant meaning, and 20 ms for targets related to the subordinate meanings).

Discussion

The results of Experiments 1-a and 1-b suggest that meanings of isolated heterophonic homographs were retrieved as predicted by an ordered-access model. The meaning of the dominant phonological alternative was accessed faster than that of the subordinate phonological alternative. However, the time course of activating the subordinate meanings was different from that found with English homophonic homographs (Simpson & Burgess, 1985) in several ways. The subordinate meanings in Simpson and Burgess's study have been already activated at 100 ms, and decayed after 300 ms from stimulus onset. In contrast, the meanings of subordinate phonological alternatives in the present study was not available at 100 ms.

The subordinate alternatives were active at 250 ms and, in contrast to English, they were still available as late as 750 ms from stimulus onset. Hence, the present data suggest that subordinate meanings of heterophonic homographs are accessed slower than the subordinate meaning of polysemous words, but they remain active for a longer time.

The divergence between the time course of disambiguating Hebrew heterophonic homographs and English homophonic homographs might reflect language-related differences or, alternatively, basic differences in processing heterophonic and homophonic homographs. However, before going any further in speculating about mechanisms of

disambiguation of homographs, it was important to make sure that the dominant and subordinate forms of the present stimuli were equivalent in their efficiency to prime their respective targets. To control for differences in accessing dominant and subordinate meanings in absence of phonological ambiguity, and to understand better the independent relationship between the dominant and the subordinate phonological alternatives of one letter string and their respective meanings, a second experiment was conducted. In the second experiment we examined the pattern of semantic facilitation of targets related to each meaning, when the phonological units to which they were related were presented in a disambiguated form.

EXPERIMENTS 2-A AND 2-B

The interpretation of the apparently ordered retrieval of the subordinate and the dominant meanings of the phonologically ambiguous letter strings presented in Experiments 1-a and 1-b was based on the relative magnitude of priming effects. This interpretation assumed that the observed difference between dominant and subordinate meanings of the primes is accounted for by their phonological ambiguity. In other words, it was assumed that in a disambiguated form, the subordinate and the dominant primes would have primed their respective targets equally. The purpose of Experiment 2 was to test this assumption.

Hebrew provides a unique opportunity to compare semantic priming effects involving alternative meanings of homographs with the semantic priming effects involving the same words presented explicitly, i.e., in a non-ambiguous form. In contrast to homophonic homographs that can be disambiguated only by semantic context (for example by embedding the homograph in a sentence), Hebrew heterophonic homographs can be disambiguated and still be presented as isolated words. This can be achieved by adding the diacritical dots to the ambiguous letter strings. The advantage of this procedure is that the experimental structure and the priming conditions remain constant for the ambiguous and unambiguous presentations.

Method

Subjects. Eighty undergraduate students, all native Hebrew speakers, participated for course credit or for payment. None of the subjects participated in the previous experiments. As in the previous experiments, 40 subjects were tested with prime/target SOAs of 100 ms and 250 ms (Experiment 2-a), and the other 40 with 750 ms SOA (Experiment 2-b).

Stimuli, design, and procedure. The stimuli, experimental design, and procedure were identical to those used in Experiments 1-a and 1-b, except that all the words and nonwords were presented in conjunction with vowel marks. Thus, each word was presented in an unequivocal phonological form, and had only one meaning.

Results

At all SOAs and with both dominant and subordinate primes, RTs to related targets were faster than RTs to unrelated targets (Table 2).

The statistical significance of the priming effects at 100 ms and 250 ms SOAs in Experiment 2-a was assessed by ANOVA across subjects (F1) and across stimuli ($F2$). The main factors were semantic relatedness (related, unrelated), dominance of prime (dominant, subordinate), and SOA (100 ms, 250 ms).

Table 2. *Reaction times and (percentage of errors) to related and unrelated targets in the different experimental conditions with phonologically unambiguous (voweled) primes (Experiments 2-a and 2-b).*

	Dominant Primes			Subordinate Primes			Nonwords	
SOA	**100**	**250**	**750**	**100**	**250**	**750**	**2-a**	**2-b**
Unrelated	722 (8%)	681 (10%)	716 (7%)	746 (9%)	702 (12%)	725 (8%)	767 (9%)	765 8%)
Related	690 (8%)	634 (8%)	672 (8%)	703 (8%)	664 (8%)	683 (6%)		
Priming Effect	+32	+47	+44	+43	+38	+42		

The ANOVA showed that across SOAs, RTs to targets in the related condition were faster than in the unrelated condition [$F1(1,39)=68.8$, MSe=1852, $p<.001$; $F2(1,39)=55.5$, MSe=2194, $p<.001$], RTs to targets related to dominant primes were faster than to targets related to subordinate primes [$F1(1,39)=18.6$, MSe=2013, $p<.001$; $F2(1,39)=6.3$, MSe=5491, $p<.01$], and RTs were faster at 250 ms SOA than at 100 ms SOA [$F1(1,39)=37.9$, MSe=4223, $p<.001$; $F2(1,39)=158$, MSe=1123, $p<.001$]. However, in contrast to Experiment 1-a, none of the interactions were statistically significant ($F1$, $F2<1.0$; for Relatedness by Frequency; ($F1<1.0$, $F2=1.3$ for Relatedness by SOA; $F1$, $F2<1.0$; for Frequency by SOA, and $F1=1.2$, $F2=1.0$, for the three-way interaction).

The analysis of the priming effects at 750 ms SOA in Experiment 2-b, revealed a significant effect of Semantic relatedness [$F1(1,39)=44.3$, MSe=1689, $p<.001$, $F2(1,39)=50.1$, MSe=1539, $p<.001$), and no main effect of Frequency of the prime [$F1(1,39)=2.8$, MSe=1432, $p>.09$; $F2(1,39)=1.4$, MSe=2692, $p>.19$). The interaction between the two factors was not significant ($F1$, $F2<1.0$). The effects of semantic facilitation obtained with the fillers at 250 ms SOA in Experiment 2-b, were very similar to the effects obtained with targets at the same SOA in Experiment 2-a (49 ms for targets related to the dominant alternatives, and 47 ms for targets related to the subordinate alternatives).

Discussion

The absence of an interaction between semantic priming and the frequency of the prime revealed that, in disambiguated form, the dominant and the subordinate phonological alternatives of the heterophonic homographs were equally effective in facilitating lexical decisions to related targets. In addition, the results of Experiments 2-a and 2-b showed that the time course of processing high- and low-frequency unambiguous Hebrew words was similar. Hence, Experiments 2-a and 2-b suggest that the difference in processing dominant and subordinate alternative meanings of heterophonic homographs observed in Experiments 1-a and 1-b was, indeed, caused by the ambiguous nature of the primes that were both phonologically and semantically equivocal.

Although the primes in Experiments 2-a and 2-b were unambiguous, an effect of dominance was obtained. Targets related to the dominant phonological alternatives incurred faster RTs than targets related to the subordinate phonological alternatives. Because in Experiments 2-a and 2-b the primes were unequivocal, this effect should be considered as a pseudodominance effect. This outcome might have resulted from our design in which different targets followed identical ambiguous primes. Consequently, the comparison across dominant and subordinate categories involved different target words. It is possible that there were intrinsic decision time differences between the target words, such that targets that happened to be related to the dominant alternatives were accessed faster than targets related to the subordinate alternatives. However, since the conclusions concerning semantic facilitation depend on the interaction *within* prime categories (comparing RTs to the same target in related vs. unrelated conditions), the pseudodominance effect has no theoretical importance.

In Experiments 3-a and 3-b we sought to examine the possible sources of the differences between the time course of activation found with English homophonic homographs (e.g., Simpson & Burgess, 1985), and between our present results with Hebrew heterophonic homographs. We endeavored to isolate the effects of semantic and phonologic ambiguity and to control for possible language specific factors. For this purpose, we have used the design of Experiment 1 with a new set of stimuli. These were Hebrew homophonic homographs, i.e., words like "BANK," that have two meanings but only one pronunciation.

EXPERIMENTS 3-A AND 3-B

Experiments 3-a and 3-b examined the time course of activation of dominant and subordinate meanings of Hebrew homophonic homographs. Each stimulus was a pattern of letters representing only one word (one phonological unit); that word, however, had two meanings, one more frequent than the other. Consequently, like most English homographs, these stimuli were semantically ambiguous but phonologically unequivocal. Using exactly the same design as in the previous experiments, the present experiments allowed comparison of homophonic and heterophonic homographs within one language - Hebrew.

Method

Subjects. The subjects were 120 undergraduates, native Hebrew speakers. They participated in the experiments for credits or payment. Sixty subjects participated in Experiment 3-a, and 60 participated in Experiment 3-b.

Stimuli. The primes were 36 ambiguous homophonic homographs that were selected from

a pool of 120 homographs. Each selected word had a dominant and a subordinate meaning. Dominance was determined empirically by the following procedure: 50 subjects rated the frequency of the meanings of all homographs on a 7-point scale from very infrequent (1) to very frequent (7). Because naming could not distinguish between meanings of homophonic homographs, the rated frequencies were validated differently than in Experiment 1. A group of 32 subjects were read a list containing only homophonic homographs, the meanings of which were rated at least 1 point apart on the frequency scale. These words were read one at a time. The subjects responded verbally with their first association to each word. The meaning that the subjects had in mind was inferred from their response. Dominant meanings were those that were produced by at least 66% of the subjects, and subordinate meanings were those that were not produced by more than 33% of the subjects. Each prime was paired with two target words: One was semantically related to the dominant meaning and the other to the subordinate meaning. Thirty-six additional homophonic homographs from the same pool were used to form semantically unrelated pairs. In addition to the word-word pairs, 72 word-nonwords pairs were again introduced as fillers. The words were homophonic homographs that were taken from the original pool. The 72 nonwords were taken from Experiment 1-a.

Design and procedure. The design of Experiments 3-a and 3-b was identical to that of Experiments 1-a and 1-b. One group of 60 subjects were tested using SOAs of 100 ms and 250 ms between primes and targets. Fifteen subjects were assigned to each of four lists, structured exactly as in Experiment 1-a (except that in each list there were 9 targets rather than 10 in each condition). Across lists, each target appeared in both related and unrelated conditions, and at both SOAs.

The second group of 60 subjects was tested using the same stimulus lists, with a design identical to Experiment 1-b, that is with the longer SOA (750 ms). Although separate analyses were conducted in each group, we will report all the results in one section.

Results and Discussion.

The RTs in the related condition were faster than in the unrelated condition at all SOAs, for dominant as well as for subordinate targets (Table 3).

Separate ANOVAs were conducted to assessed the reliability of the priming effects across subjects ($F1$) and across stimuli ($F2$), at 100 ms and 250 ms SOAs (Experiment 3-a). These ANOVA showed that across SOAs, RTs to targets in the related condition were faster than in the unrelated condition [$F1(1,59)=19.7$, MSe=1957, $p<.001$; $F2(1,35)=15.6$, MSe=1690, $p<.001$], RTs to targets related to dominant primes were faster than to targets related to subordinate primes [$F1(1,59)=14.6$ MSe=1675, $p<.001$; $F2(1,35)=4.5$ MSe=5166, $p<.04$], and RTs were faster at 250 ms SOA than at 100 ms SOA [$F\ 1(1,59)=145$. MSe=2035, $p<.001$; $F2(1,35)=250$. MSe=675, $p<.001$]. As with unambiguous primes in Experiments 2-a and 2-b and in contrast to Experiments 1-a and 1-b, semantic relatedness did not reliably interact with any other factor.

Table 3. *Reaction times and (percentage of errors) to related and unrelated targets in the different experimental conditions with homophonic homographs as primes (Experiments 3-a and 3-b).*

	Dominant Primes			Subordinate Primes			Nonwords	
SOA	**100**	**250**	**750**	**100**	**250**	**750**	**3-a**	**3-b**
Unrelated	591 (6%)	545 (8%)	567 (6%)	606 (7%)	561 (8%)	580 (7%)	680 (9%)	653 (8%)
Related	580 (6%)	522 (6%)	553 (6%)	588 (6%)	540 (8%)	570 (8%)		
Priming Effect	+11	+23	+14	+18	+21	+10		

The analysis of the priming effects at 750 ms SOA (Experiment 3-b), showed that the semantic relatedness effect was reliable [$F1(1,59)=7.9$, MSe=1098, $p<.007$; $F2(1,35)=7.6$, MSe=694, $p<.009$] and RTs to targets were faster following dominant primes than following subordinate primes [$F1(1,59)=10.6$, MSe=1265, $p<.002$; $F2(1,35)=3.9$ MSe=3323, $p<0.05$]. As with the shorter SOAs, the interaction between the two factors was not reliable ($F1, F2<1.0$).

The most important finding in Experiments 3-a and 3-b was that, in absence of phonological ambiguity, both the dominant and the subordinate meanings of Hebrew polysemous words were already available at 100 ms from stimulus onset. Similarly to heterophonic homographs, they remained active at least during the first 750 ms. These results suggest that the distinct pattern of activation observed for low- frequency phonological alternatives of heterophonic homographs (in Experiment 1-a) was caused by phonological rather than semantic ambiguity.

Because our study did not include a condition of very short SOA (16 ms) between primes and targets, the onset of activating dominant and subordinate meanings of Hebrew homophonic homographs cannot be directly compared to the pattern of activation reported by Simpson and Burgess (1985) with English materials. However, the persistent activation of subordinate meanings at the longer SOA of 750 ms in the present experiment, clearly differs from the pattern of activation observed in English (Simpson & Burgess, 1985). This divergence suggests that the process of disambiguating polysemous words might involve language-specific components. Possible interpretations of these results are elaborated in the general discussion.

Across experiments comparisons

Several formal comparisons were conducted to assess priming effects involving heterophonic primes at all SOAs (Experiments 1-a and 1-b) and priming effects involving homophonic primes (Experiments 3-a and 3-b). For these analyses the relevant data from the four experiments were combined in mixed ANOVA designs in which the type of homographs was introduced as an additional between-subjects factor. First, we compared the pattern of semantic facilitation of the subordinate meanings only, across all SOAs, for the two types of homographs. The three-way interaction of relatedness, SOA, and homograph type was significant ($F1(1,98)=6.9$, MSe=1914, $p<0.009$; $F2(1,74)=3.6$, MSe=2926, $p<0.06$), suggesting a reliable difference in the time course of activating the subordinate meanings of heterophonic and homophonic homographs.

Another finding regarding the two types of homographs was that the average effects of semantic priming of the dominant alternatives across all SOAs, were twice as strong for heterophonic homographs (35 ms facilitation), than for homophonic homographs (16 ms facilitation). The statistical significance of this difference was assessed by a mixed ANOVA design in which RTs to the dominant meanings of heterophonic homographs at all three SOAs, were compared with the respective RTs to the dominant meanings of homophonic homographs. The type of homography served again as a between subjects factor. This analysis revealed a significant interaction of relatedness and homography type ($F1(1,98)= 7.6$, MSe=1675, $p<0.007$; $F2(1,74)=6.0$, MSe=1994, $p<0.02$). Whether the shrinking of the priming effect for homophonic homographs relative to heterophonic homographs reflects primarily differences in processing the two types of homographs, or merely a floor effect due to much faster responses to homophonic than heterophonic homographs, was not clear. Therefore, we replicated Experiment 1-a using an identical number of subjects and identical methods.

The purpose of replicating Experiment 1-a was, in fact, two-fold. First, because the comparison of heterophonic and homophonic homographs was based on a different pool of subjects, and because the most important difference relied on one data point, we aimed at reexamining the absence of priming effect (or the possible inhibition) for heterophonic homographs at 100 ms SOA. Second, to examine whether the larger priming effects found for heterophonic relatively to homophonic homographs were due to an incidental overall slower performance of the subjects sampled in Experiment 1-a.

The results of this replication are presented in Table 4. As in the original experiment, lexical decisions for targets related to the subordinate meanings of the primes were not facilitated at 100 ms SOA. In addition, the nonsignificant trend of inhibition observed in this condition in Experiment 1-a proved to be unreliable. Overall, the RTs in the replication were faster than the original experiment. This suggests that the subjects employed in Experiment 1-a were generally slower than all other subjects in this study.

Table 4. *Reaction times and (percentage of errors) to related and unrelated targets in the replication of Experiments 1-a.*

	Dominant Primes		Subordinate Primes		Nonwords
SOA	100	250	100	250	
Unrelated	626 (9%)	588 (10%)	635 (13%)	609 (11%)	677 (8%)
Related	601 (8%)	557 (5%)	635 (10%)	588 (10%)	
Priming Effect	+25	+31	0	+21	

Nevertheless, the pattern of the semantic facilitation with unvoweled ambiguous heterophonic homographs was replicated. The statistical significance of the priming effects was assessed by ANOVA across subjects ($F1$) and across stimuli ($F2$). The main effects of relatedness and dominance were significant [$F1(1,39)=4.6$, MSe=1436, $p<0.04$; $F2(1,39)=5.0$, MSe=1834, $p<0.03$; $F1(1,39)= 12.6$, MSe=1552, $p<0.001$; $F2\ (1,39)= 9.9$, MSe=2648, $p<0.003$; respectively]. The two-way interaction did not reach significance in the stimuli analysis [$F1(1,39)=3.6$, MSe=1741, $p<0.06$; $F2(1,39)=2.4$, MSe=1756, $p<0.1$]. Planned comparisons revealed that RTs to targets related to the subordinate alternatives of the prime-meanings at 250 ms SOA were significantly faster than RTs to targets in the unrelated condition [$t(1,39)=3.1$, $p<0.004$]. We will refer to the additional implications of this replication in the General Discussion.

GENERAL DISCUSSION

In the present study we examined the process of disambiguating Hebrew heterophonic and homophonic homographs presented in the absence of biasing context. To summarize the results of our investigation, it appears that regardless of relative dominance, at least two different meanings of each homograph were retrieved. However, the time-course of activating the different meanings and possibly the amount of activation were influenced by phonological factors. With homophonic homographs, subordinate as well as the dominant meanings were active as early as 100 ms from stimulus onset. On the other hand, with heterophonic homographs, only the dominant meaning was available at 100 ms SOA, whereas the availability of the subordinate meaning was delayed. In contrast to the differences found at the onset of meaning activation, the decay of activation of subordinate meanings of homophonic and heterophonic homographs was similar; they all remained active as late as 750 ms from stimulus onset. Thus, the onset activation pattern of Hebrew heterophonic homographs observed in the present study is in agreement with the ordered-access model suggested by Simpson and Burgess, (1985). At present this conclusion must be limited to heterophonic homographs because unlike Simpson and Burgess (1985), we did not use SOAs shorter than 100 ms. Our findings suggest then, that heterophonic homographs and homophonic homographs are disambiguated differently. This difference and the long lasting activation of subordinate meanings of Hebrew but not English homographs, may provide some insights regarding the lexical structure and the process of word identification.

The lexical representation of homophonic homographs is controversial. Some authors assert that homophonic homographs entertain different lexical entries, one for each meaning (Forster & Bednall, 1976; Jastrembski, 1981; Kellas et al., 1988). Other authors claim that a homograph has only one lexical entry, related to multiple nodes in a semantic network (Seidenberg et al., 1982; Cottrell & Small, 1983). On the other hand, heterophonic homographs are, by definition, represented by several phonological units in the lexicon. Thus, phonologically ambiguous letter strings refer to different lexical entries, one for each phonological realization. The relatively delayed access to the subordinate meanings of heterophonic homographs, as compared with subordinate meanings of homophonic homographs could be more easily accounted for by assuming only one lexical entry for homophonic homographs and several entries for heterophonic homographs.[3]

According to such a model, the alternative lexical entries are automatically activated by the unique orthographical pattern, though at different onset times. The present data and the results of our previous studies (e.g., Bentin & Frost, 1897) indicate that, in the absence of biasing context the order of activation is determined by the relative word frequency; higher-frequency words are accessed before lower frequency words. As a consequence of the multiple entries structure and the ordered-access process, heterophonic homographs are phonologically disambiguated *before* the semantic network is accessed. Each activated word (in the lexicon) is unequivocally related to a meaning. Because entries of dominant words are accessed before those of subordinate words, the origin of the dominance effect on the time course of activating the meanings of a heterophonic homograph could have been the well documented frequency effect on lexical access.

This interpretation may also account for the overall greater priming effects found for heterophonic than for homophonic homographs. It might suggests that when one lexical unit activates two or more semantic nodes, each of these nodes is activated less than nodes which are unequivocally related to phonological units in the lexicon. If, in contrast to heterophonic homographs, homophonic homographs were represented by only one lexical entry which is related to several semantic nodes, the process of disambiguating the different meanings should have been less affected by the relative frequency (dominance) of using each meaning. Our hypothesis is that activating a lexical entry in an unbiased semantic context, should automatically initiate the retrieval of all its related meanings. Because only one lexical entry is active the relative dominance of the alternative meanings is irrelevant at the stage of lexical access. Relative frequency factors might affect the order of their retrieval at later processing stages, but our results suggest that, at least for the SOAs that have been examined in the present study, such an effect was not observed.

One caveat that must be considered while interpreting the difference in the amount of priming with homophonic vs. heterophonic homographs is that the former were overall faster than the latter. The reduction in overall RTs latencies in the replication of Experiment 1-a relative to the original experiment, and the comparison of the nonword data across all experiments help to clarify this issue. Because the RTs to nonwords in the replication were identical to those in Experiment 3-a, we can assume that these two groups of subjects were comparable in overall speed of performance. Nevertheless, RTs to targets related to heterophonic homographs were slower by about 40 ms than RTs to targets related to homophonic homographs. This difference was not entirely unexpected; it conforms with previous finding in Hebrew showing faster RTs for phonological unequivocal words than for phonologically ambiguous words (Bentin et al. 1984). However, this pattern might have caused a floor effect in the RTs to homophonic homographs that attenuated the absolute magnitude of the priming effect. A floor effect as a sole explanation for this attenuation is not entirely supported by the data. Note, that although the overall difference in RTs in the two homographic conditions was reduced by a factor of 3 (from 120 to 40 ms) when the replication rather than the original Experiment was considered, the respective reduction of the priming effect for the dominant alternatives was relatively small (from 35 ms to 28 ms). Second, the smallest effect of semantic facilitation for homophonic homographs (11 ms) was obtained for dominant primes at 100 ms SOA that were *slower* by 50 ms than the primes at 250 ms SOA (which revealed a much larger facilitation). Thus, the observed difference in the magnitude of the priming effects between homophonic and heterophonic homographs is consistent with the hypothesis that they have different lexical structures.

In addition to the implications on the lexical structure, our data are also relevant to arguments regarding the use of phonology in word identification. One class of models suggest that (with the possible exception of very infrequent words), printed words activate orthographic units that are directly related to meanings in semantic memory (e.g., Seidenberg, 1985; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Such a mechanism had been invoked to explain, for example, how homophones such as SALE and SAIL are correctly understood, or how patients with acquired dyslexia can understand written words without being able to read them aloud (Kay & Patterson, 1985).

An alternative class of models asserts that, most of the time, access to meaning is mediated by phonology (e.g., Perfetti, Bell, & Delaney, 1988; Van Orden, Johnston, & Hale, 1988). The latter class of models is supported by theoretical considerations such as the parsimony of having only one mechanism that mediates access to meaning for both speech and reading (e.g.,

Liberman & Mattingly, 1989), and by evidence that when the ability to derive phonology from print is poor (as in deep dyslexia), semantic errors in reading are abundant (see for a review Marshall & Newcombe, 1980). Moreover, when the direct connection from orthography to meaning is impaired (as in some patients with surface dyslexia), the meaning of printed words can be retrieved by pre-lexical application of grapheme-to-phoneme (GTP) transformation rules (Coltheart, Masterson, Byng, Prior, & Riddoch, 1983; Marshall & Newcombe, 1973).

As we have pointed out in previous papers, lexical decisions for unvoweled Hebrew words are based primarily on orthographic codes (Bentin et al., 1984; Bentin & Frost, 1987), and even for naming, pre- lexical word phonology is not usually used by skilled readers (Frost et al., 1987). Nevertheless, the present results suggest that, in contrast to lexical decisions, the retrieval of meaning requires the activation of the phonological structure to which the printed word refers. If meaning were retrieved directly from the orthographic input, no difference should have been found between processing homophonic and heterophonic homographs. The delayed onset of activating the meanings of the subordinate phonological alternatives relative to the subordinate meanings of homophonic homographs, and possibly the overall more robust priming effects observed when the primes were phonologically ambiguous than when they were homophonic homographs, suggests that the former involved phonological disambiguation prior to the disambiguation of meaning.

One of the most intriguing results of the present study was that subordinate meanings of both heterophonic and homophonic homographs were still available and used 750 ms from stimulus onset. This result contrasts with the relatively fast decay of subordinate meanings of English homographs (Simpson & Burgess, 1985; Kellas et al., 1988). Because the decay pattern was similar for both types of Hebrew homographs, the divergence from English should be probably accounted for by language-related factors. One possible source of the different results obtained in Hebrew and in English may be related to the homographic characteristics of the Hebrew orthography. Hebrew, like other Semitic languages, is based on word families derived from tri-consonant roots. The root is contained in all of its derivations, therefore, Hebrew contains many homophonic and heterophonic homographs. The wide spread of homography might have shaped the reader's reading strategies. Because ambiguity is so prevalent in reading, the process of semantic and phonologic disambiguation is governed mainly by context. As the disambiguating context often follows rather than precedes the ambiguous homographs, the most efficient strategy of processing them should consist of maintaining their phonologic or semantic alternatives in working memory until the context selects the appropriate one. Note that by this interpretation the subordinate alternatives do not decay automatically, but remain in memory until disambiguation by context has occurred. However, a complete account of the specific characteristics of the Hebrew orthography which might have influenced our results with homophonic and heterophonic homographs deserves further investigation.

REFERENCES

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning Memory & Cognition, 10,* 353-368.

Bentin, S., & Frost, R. (1987). Processing lexical ambiguity and visual word recognition in a deep orthography. *Memory & Cognition, 25,* 13-23.

Carpenter, P. A., & Daneman, M. (1981). Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior, 20,* 137-160.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12,* 590-606.

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *Quarterly Journal of Experimental Psychology, 35A,* 469-495.

Cottrell, G. W., & Small, S. L. (1983). A connectionist scheme for modeling word sense disambiguation. *Cognition and Brain Theory, 6,* 89-120.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation time in reading. *Journal of Memory and Language, 27,* 429-446.

Forster, K. I., & Bednall, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition, 4,* 53-61.

Frost, R., Feldman, L. B., & Katz, L. (1990). Phonological ambiguity and lexical ambiguity: Effects on visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 569-580.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104-114.

Glucksberg, S., Kreuz, R. J., & Rho, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 323-335.

Hogaboam, T. W., & Perfetti, C.A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 14,* 265-274.

Holley-Wilcox, P., & Blank, M. A. (1980). Evidence for multiple access in the processing of isolated words. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 75-84.

Jastrembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13*, 278-305.

Kay, J. K., & Patterson, K. E. (1985). Routes to meaning in surface dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neurophysiology and cognitive studies of phonological reading* (pp. 79-104). London: Lawrence Erlbaum Ltd.

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the time course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 601-609.

Kroll, J. F., & Schweickert, J. M. (1978). Syntactic disambiguation of homographs. Paper presented at the Nineteenth Annual Meeting of the Psychonomic Society, San Antonio, Texas.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Marshall, J. C., & Newcombe F. (1973). Patterns of paralexia. *Journal of Psycholinguistic Research, 2*, 175-199.

Marshall, J. C., & Newcombe F. (1980). The conceptual status of deep dyslexia: An historical perspective. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia*. London: Routlage & Kegan Paul.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Neil, W. T., Hilliard, D. V., & Cooper, E. (1988). The detection of lexical ambiguity: Evidence for context-sensitive parallel access. *Journal of Memory and Language, 27*, 279-287.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9*, 225-236.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 779-790.

Seidenberg, M. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowsky, M. (1982). Automatic access of the meaning of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537.

Seidenberg, M. S., Waters, G. S., Barnes, M., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Schvaneveldt, R. W., Meyer, D. E. & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology :uman Perception & Performance, 2*, 243-256.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior, 20*, 120-136.

Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin, 96*, 316-340.

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception & Performance, 11*, 28-39

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-660.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440.

Van Orden, G. C., Johnston, J. C., and Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related Potential analysis of the time course of meaning activation. *Journal of Memory and Language, 26*, 188-208.

FOOTNOTES

Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 58-68.

†Department of Psychology, The Hebrew University, Jerusalem.

[1]Simpson and Burgess (1985) found no difference in RTs between SOAs of 100 and 300 ms. However, throughout the present study, the main effect of SOA was quite reliable and robust (including in the replication of Experiment 1-a), suggesting that unlike S and B, in the present study the 100 ms SOA condition was more difficult than the other SOA conditions. A possible explanation of this difference is that our procedure did not include an initial fixation point.

[2]Although phonologic ambiguity is very common in the Hebrew orthography, the set of stimuli used in the experiments was constrained by many experimental controls such as mean rated frequencies, dominance as reflected by naming performance, syntactic classes, rated semantic relatedness, etc. This set of stimuli did not permit a within-subject design across all SOAs. A similar problem was raised and solved similarly by Simpson and Burgess 1985).

[3]Seidenberg et al. (1982) present a similar kind of single vs. multiple argument for noun-noun vs. noun-verb homophonic homographs, but they draw slightly different inferences. They argue that noun-verb homographs (e.g., train) have different entries in the lexicon and, hence, both meanings are always accessed for such words even when a strong priming word is presented in the context. In contrast, for noun-noun ambiguities (e.g., boxer) there is only one entry in the lexicon, and meanings are accessed in order of relative activation levels.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 205-220

Attention Mechanisms Mediate the Syntactic Priming Effect in Auditory Word Identification*

Avital Deutsch† and Shlomo Bentin†

The effect of syntactic priming and the involvement of attention in that process was investigated testing identification of spoken Hebrew words presented in sentences. Target words were masked by white noise and were either congruent or incongruent with the syntactic structure of the sentence. In comparison to a neutral condition, the identification of congruent targets was facilitated and identification of incongruent targets was inhibited, equally. When congruent and incongruent sentences were presented in separate blocks the inhibition effect was attenuated whereas the facilitation was not affected. The introduction of 350 ms silent ISI between the context and the target increased the inhibition without affecting the facilitation. We suggest that the facilitation as well as the inhibition effects of syntactic priming are based on a veiled controlled process of generating expectations. The inhibition is caused by an additional controlled process of re-evaluation of the auditory input triggered by syntactic incoherence. The later process requires additional attentional resources.

There is much evidence in the research literature that syntactic context influences the process of word recognition (Carrello, Lukatela, & Turvey, 1988; Goodman, McClelland, & Gibbs, 1981; Gurjanov, Lukatela, Moskovljević, & Turvey, 1985; Katz, Boyce, Goldstein, & Lukatela, 1987; Lukatela, Kostić, Feldman, & Turvey, 1983; Lukatela & Moraco, Stojonov, Savić, Katz, & Turvey, 1982; Marslen-Wilson, 1987; Seidenberg, Waters, Sanders, & Langer, 1984; Tanenhaus, Leiman, & Seidenberg, 1979; Tyler & Wessels, 1983; West & Stanovich, 1986; Wright & Garret, 1984). The common finding is that performance is faster and more accurate if the target words are congruent with the syntactic structure into which they are integrated, than when they are incongruent. This differential performance was found mostly in tasks such as lexical decision and naming (Carello et al., 1988; Goodman et al., 1981; Katz et al., 1987; Gurjanov et al., 1985; & Lukatela et al., 1982; 1983; Seidenberg et al., 1984; Tanenhaus et al., 1979; West & Stanovich, 1986; Wright & Garrett, 1984). In analogy with the effects of semantic context in similar tasks, the influence of the syntactic context has often been labeled grammatical or syntactic priming. However, because the term priming has been borrowed from the semantic domain, the use of the term "priming" in the syntactic domain needs specific consideration. In the semantic domain, priming refers primarily to a process that influences the identification of a particular lexical entry (Forster, 1981; Seidenberg, 1982). The syntactic context, on the other hand, refers primarily to a particular grammatical form of the word, which may or may not be independently represented in the lexicon. Therefore syntactic priming may affect the identification of a particular grammatical structure, without direct influence on accessing a particular lexical entry. It is in this sense that we will adopt here the term syntactic priming.

This study was supported by a grant from the Israeli Foundation Trustees to Shlomo Bentin and by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. Avital Deutsch was supported by a stipend from the Martin and Vivian Levin Center. We thank Len Katz, and Don Shankweiler for useful comments on an earlier draft. We wish to express special appreciation of Ram Frost's continuous help and constructive criticism.

Like other priming phenomena, syntactic priming might also reflect the combined or independent contribution of two basic components: One is the facilitation of processing syntactically congruent targets due to the agreement between the observed grammatical form and that predicted by the syntactic structure. The other is the inhibition of processing incongruent targets either because they do not conform with previous expectation, or because they may require additional processing aimed at resolving the amorphic input, or both. Several studies have interpreted syntactic priming in terms of facilitation (Katz et al., 1987; Lukatela et al., 1982; 1983; Marslen-Wilson, 1987; Tyler & Wessels, 1983), while others have emphasized the inhibitory aspect (Tanenhaus et al., 1979; West & Stanovich, 1986; Carello et al., 1988). However, the question of whether facilitation or inhibition, or both are operative is unsettled because, with the exception of one study in which only inhibition was found (West & Stanovich, 1986), the syntactic priming effect has not been assessed relative to a neutral condition.

The distinction between facilitation and inhibition is important because each of these two processes might reflect a different cognitive mechanism. In particular, current models of priming suggest that facilitation and inhibition differ in their attentional requirements. In normal language communication syntactic congruity is expected. Therefore, it might be expected that syntactically congruent targets are automatically integrated into the sentence structure. In contrast, syntactically incongruent targets cannot be automatically integrated into the syntactic context. Therefore, they may require some re-evaluation of the sensory input as well as of the context. In the semantic domain it is assumed that these activities which inhibit word identification are actively controlled and require the allocation of attention resources (Neely, 1977; Posner & Snyder, 1975).

The role of attention in syntactic priming has been approached indirectly in earlier studies. For example, dealing with inflectional morphology, Katz et al. (1987), suggested a modular syntactic processor whose involvement in word recognition is mandatory and informational encapsulated (Fodor, 1983). This interpretation implies that syntactic priming, particularly as it relates to facilitatory processes, should not require attentional resources. Indeed, several authors have proposed that syntactic priming is automatic (Carello et al., 1988; Gurjanov et al., 1985; Lukatela et al., 1982: See also Seidenberg et al., 1984)). Note, however, that in the studies just cited, the automaticity of the syntactic priming effect was suggested primarily by inflectional processing in pairs of words presented in the highly-inflected Serbo-Croatian language. Testing English-speaking subjects with word-pairs materials, Goodman et al. (1981) found evidence that syntactic priming may be strategy-controlled and modulated by attention. A role for attention in syntactic priming can be inferred indirectly from the assumption that attention is involved primarily during lexical (or post-lexical) processes which are involved in lexical decision more than in naming. Indeed, several studies using single-word context in English (Seidenberg et al., 1984) as well as in Serbo-Croatian (Carello et al., 1988) reported that syntactic priming in naming, was significantly smaller than in lexical decision or inexistent. In addition, using sentential context West and Stanovich (1986) found significant inhibition for incongruent targets without facilitation of congruent targets.

The involvement of attention may be especially conspicuous in the case of incongruent targets, when re-evaluation of the target/sentence relationship, although possibly unavoidable, necessarily requires attentional resources. Such an interpretation of the syntactic priming effect was suggested by Tanenhaus et al. (1979). Examining the process of selecting the contextually appropriate readings of noun-verb ambiguities in sentences, these authors suggested that the syntactic selection process is characterized by veiled controlled mechanism which makes use of context to suppress the inappropriate meaning (see Shiffrin & Snyder, 1977). Applied to syntactic priming, Tanenhaus et al. (1979), suggest that the inhibition of incongruent targets is caused by a controlled, yet unavoidable (therefore "veiled") process of matching the incongruent sensory input with the expected syntactic structure.

To summarize, the present evidence for a role of attention in syntactic processing is not conclusive. Indeed, most authors suggest that the application of syntactic rules is mandatory and does not require much attention. However, the empirical basis for this conclusion is weak. First, attention was not directly manipulated in any of those studies. Second, the conclusions were based mostly on studies of syntactic priming by single word context. Finally, the absence of neutral conditions in most studies prevents any distinction between the facilitatory and inhibitory

components of the syntactic priming effect. The present study is a systematic investigation of the syntactic priming effect in spoken sentences. We sought to determine the relative contribution of facilitatory and inhibitory mechanisms to syntactic priming and to examine the attention requirements of each of these mechanisms.

Methodological Considerations

In the present study, we manipulated the Hebrew agreement rule between subject and predicate regarding gender and number, and a morpho-syntactic rule that involves the decomposition of the conjunctive form of pronoun-plus-preposition. These rules were chosen for two reasons. First, we aimed at isolating the influence of the syntactic context from the influence of the semantic context. Both agreement between subject and predicate, and the morpho-syntactic rule that we employed are simple and essential in Hebrew grammar. The essential role of an agreement rule in Hebrew is to specify the syntactic relation between the constituents of a sentence, and has no effect on the semantic information. For example: The predicate agrees with the subject in person, gender and number but, because the specification of the gender and number is already available in the subject, violation of one or more of these types of agreement does not affect the meaning of the sentence. Moreover, because the agreement rule is at the level of inflectional morphology, violation of it does not cause changes in word class (changes, that may have semantic implications, Carello, 1988).

Second, the particular agreement rule that we chose operates between sentential elements, like the subject and the predicate, and not at the phrase level as, for example, the agreement between subject and attribute. Therefore, we were not constrained to present the subject and the predicate in succession, thus emphasizing the sentence rather than the phrase level. Because of the minimal involvement of semantic factors, and the possibility to deal with syntactic rules beyond the phrase level, we believed the rules that we used, were appropriate for exploring syntactic priming effect.[1] In addition it should be emphasized that none of the targets used represented a high cloze of the sentence. Therefore, subjects could not simply predict the target and use semantically-induced word-guessing strategy.

Most of the previous studies of the effect of syntactic context (with the exception of Katz et al., 1987; Marslen-Wilson, 1987 and Tyler & Wessels, 1983) used visually presented stimuli. In the present study, we have examined syntactic priming in speech perception rather than reading because speech is more basic than reading in human language and is perhaps less affected by learned strategies.

Previous studies of semantic or associative priming in the visual modality suggested that the degradation of stimulus intelligibility magnifies the effect of contextual influence on word recognition (Becker & Killion, 1977; Meyer, Schvaneveldt, & Ruddy, 1975; Neely, 1991; Stanovich & West, 1983). Therefore, in attempt to focus our investigation on the nature of the syntactic context effect, our basic task required the identification of target words masked by white noise.

EXPERIMENT 1

The purpose of the present experiment was to assess the relative contribution of facilitatory and inhibitory processes to syntactic priming. In a previous study (Bentin, Deutsch, & Liberman, 1990) we observed a large syntactic context effect on the identification of words masked white noise. The identification of target words was four times as accurate when they were syntactically congruent than when they were incongruent with the context sentence. In the present experiment we replicated and extended our former study by adding a neutral condition. The addition of the neutral condition enabled us to disentangle the facilitatory effect of syntactic congruity and the inhibitory effect of syntactic incongruity that were confounded in our previous study (see also West & Stanovich, 1986, Neely, 1976).

The neutral context that we used with all targets was "the next word is...," as was originally suggested by McClleland and O'Regan (1981) and applied to an investigation of syntactic priming in reading by West and Stanovich (1986). We chose this neutral condition because, it probably involves no syntactic bias toward specific syntactic structures or word classes (West & Stanovich, 1986).

We assumed that the facilitatory and inhibitory components which may contribute to the syntactic priming effect, should be differentially reflected in comparison to the neutral condition. Facilitation was measured by the difference between the percentage of correct target identification in the congruent and the neutral context, whereas the difference between the correct identification in the neutral and the incongruent context conditions was the measure of inhibition.

Method

Subjects. The subjects were 30 undergraduate students who participated in the experiment for course credit or for payment. They were all native speakers of Hebrew, without any known hearing problems.

Test Materials. The auditory identification test included 44 target words. Targets were the last word in a three- or four-word sentence. Each target was embedded in three different sentences, which defined three different conditions of the syntactic context: 1. "Congruent"—the target word fitted the syntactic structure of the sentence. 2. "Incongruent"—the target word did not fit the syntactic structure of the sentence, that is, caused a violation of a syntactic rule. 3. A "Neutral" condition as explained above.

The syntactic violations were constructed by changing the congruent sentences in one of the following ways.

Type 1: Violation of the agreement in gender between subject and predicate. This category included 12 target words repeated across the three context conditions forming a total of 36 sentences. In the incongruent condition a masculine subject was presented with a feminine predicate (in 6 of the sentences) or vice-versa (in the other 6 sentences), that is, a feminine subject presented with a masculine predicate.

Type 2: Violation of the agreement in number between subject and predicate. Twelve target words (other than in Type 1) were repeated across the three conditions forming 36 sentences. In the incongruent condition a singular predicate followed a subject in a plural form (in 6 of the sentences), or vice-versa (in the other 6 sentences).

Type 3: Violation of the agreement in both gender and number between subject and predicate. This category also included 12 target words (different than in type 1 and 2) and repeated across conditions to form 36 sentences. In the incongruent condition the compatibility of gender and number between the subject and predicate was altered in each sentence. For example: A masculine singular subject was followed by a feminine plural predicate. (We constructed all the 4 possible combinations, 3 sentences for each).

Type 4: Decomposition of the conjunctive form of pronoun and preposition. This category included 8 target pronouns, each of which was combined with a different preposition, forming 24 sentences. In Hebrew, the pronoun and the preposition are always in a conjunctive form. Thus, in the incongruent condition, the conjunctive form was decomposed into its two elements. For example: The conjunctive form "alecha" ("on you") was presented as two separate words: "al" (the preposition "on") and "ata" (the pronoun "you"). In the neutral condition the targets were presented as normal conjunctions.

The sentences of types 1 to 3 were formed of three words in the following order: Subject, attribute and predicate. The masked target was always the predicate. The predicate was either a verb or an adjective (participle form in nominal clauses). Type 4 sentences were formed of a subject, a predicate and a verbal completion (the conjunctive pronoun). The masked targets were the verbal completions in their normal conjunctive form (congruent and neutral conditions) or decomposed (the incongruent condition).

The sentences were organized in 3 lists of 60 sentences, 20 in each congruity condition. Each group of 20 included 12 manipulations of the agreement rule (Types 1 to 3) and 8 manipulations of the morpho-syntactic rule (Type 4). The targets in sentences of Types 1 to 3 were rotated so that each subject saw each target only once but, across subjects, each target appeared in each congruity condition. Because the number of the pronouns is limited, the rotation of pronouns between congruity conditions was within subjects, so that each appeared 3 times in a list (once in the decomposed form). In order to avoid the repetition of priming as much as possible a different context was used in each condition. Moreover, the contexts were counterbalanced across the three lists.

All the sentences were recorded on tape by a female who was a professional speaker of Hebrew. The tapes were digitized at 20 kHz and edited as follows. The duration of the mask was equal in all sentences, determined by the duration of longest target. The white noise was digitally added to the target, starting slightly before onset with a signal-to-noise ratio of 1:3.4. This ratio was determined on the basis of pilot tests, so that correct target identification level was about 50%.

The sentences in each list were randomized and output to tape at a 2 second inter-sentence interval at a comfortable loudness.

Procedure. Subjects were randomly assigned to one of the three stimuli lists. Each subject was tested individually. The experimenter and the subject listened to the stimuli simultaneously, both using earphones (HD-420).

The subject was instructed to listen to the sentence and to repeat the last (masked) word

during the silence interval at the end of each sentence. No time constraints were imposed; in a few instances when the subject's response was delayed relative to the inter-sentence interval, the experimenter stopped the tape-recorder. The responses were recorded manually by the experimenter.

The experimental session began with 12 practice trials (4 sentences in each condition), followed by the test list.

Results

Subjects' responses were initially coded as correct (accurate identification of the inflected word) or error. The errors made in the incongruent condition were further categorized into four types: 1) "Self correction" (a correction of the syntactic violation using the same root); 2) "Random completion" (a totally different root forming a semantically and syntactically congruent sentence); 3) "Nonsense" (any completion which was semantically meaningless or syntactically incongruent, including nonwords); 4) "No response" ("I don't know"). In the neutral and congruent conditions only the last three categories were possible.

Because in our previous study (Bentin et al., 1990) the congruity effect on the four types of syntactic rules was similar, we collapsed our analysis over the sentence types.

Across subjects or stimuli, the percentages of correct identification were 74.8%, 50.2%, and 27.3% for the congruent, neutral, and incongruent syntactic conditions, respectively (Figure 1).

The statistical significance of the congruity effect was examined by one-factor analyses for subjects (F1) and stimuli (F2). The main effect of syntactic context was significant [$F1(2,58)=110.5$, $MSe=153$, $p<.0001$ and $F2(2,118)=49.8$, $MSe=661$, $p<.0001$].

The distribution of errors is presented in Table 1. Statistical evaluation of the distributions (ANOVA followed by Tukey-A post-hoc comparison) showed that within each congruity condition, all differences were reliable at the $p<.05$ level.

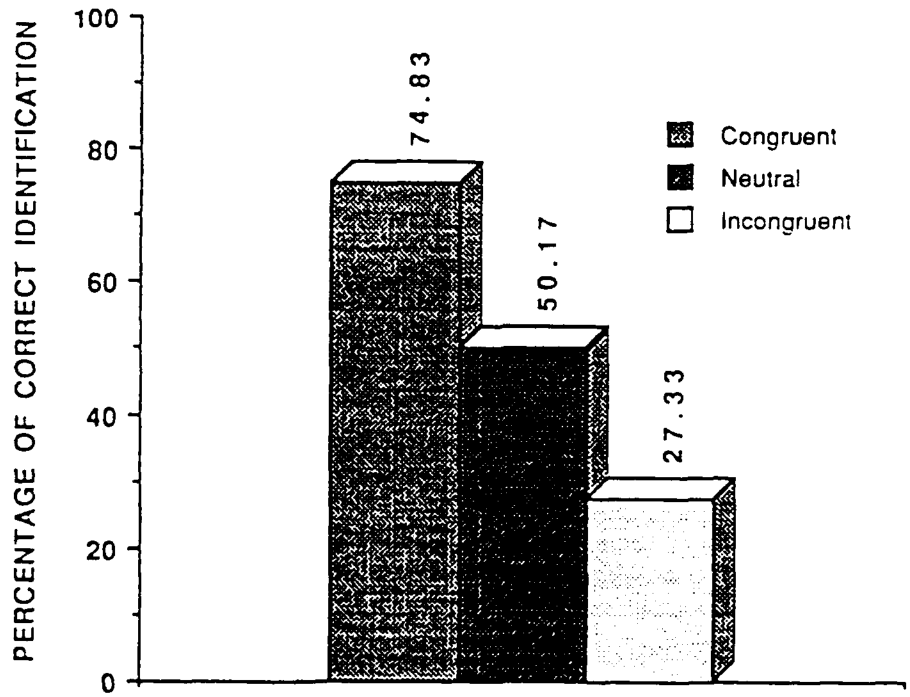

Figure 1. **The percentage of correctly identified congruent, neutral and incongruent targets.**

Table 1. *Mean percentage of errors (SEm) of each type in each congruity condition.*

	ERROR TYPE							
CONGRUITY CONDITION	SELF CORRECTION		RANDOM COMPLETION		NONSENSE		NO RESPONSE	
CONGRUENT	–		62.0%	(4.1)	2.9%	(1.9)	33.4%	(4.3)
NEUTRAL	–		54.6%	(4.0)	0.5%	0.5)	42.8%	(3.9)
INCONGRUENT	12.1%	(1.2)	19.3%	(2.1)	3.5%	(1.2)	62.2%	(3.2)

Discussion

The results of Experiment 1 demonstrated that the syntactic priming effect, as it is revealed in our auditory word identification paradigm, consists of two components, facilitation and inhibition. The relative contribution of each component to the global context effect is approximately equal: Congruent context improved identification of white-noise masked words by about 23% while incongruous context reduced identification by the same amount, from a neutral baseline of about 50% correct.

Before discussing these results any further, a trivial interpretation should be considered. Because only verbatim accurate responses were considered correct, it could have been the case that the pattern of facilitation and inhibition simply reflected that, facing uncertainty, subjects identified the word-root and completed the inflection using an intelligent-guessing strategy. Along with this interpretation, the difference in the percentage of correct identification of inflected targets in the congruent and incongruent conditions, would reflect the correspondence or disagreement between the subject's intuition about how the identified root should have been inflected and what was actually presented. Such a strategy, however, implies that a) targets' roots were identified better than their inflected forms and b) that in the incongruent condition there would have been be a high percentage of Type 1 errors (i.e., errors reflecting the inadequate use of the correct syntactic form). The first implication could not hold in the present experiment because, as mentioned in the methodological considerations, there was no strong semantic constrain which could have facilitated an independent identification of roots on semantic basis. The second was rejected by the analysis of errors.

As revealed in Table 1, the percentage of self correction in the incongruent condition was very small, by far smaller than the percentage of no responses. Note also that the percentage of random completions (i.e., substituting the target with an incorrect but semantically and syntactically congruent word) was also relatively low in this condition. This pattern does not support the "intelligent-guessing strategy" while suggesting that the low percentage of correct identification in this condition reflected a general process of inhibition caused by syntactic incongruence.

Additional support to our interpretation is provided by comparing the pattern of random completions and no responses in the incongruent condition with those observed in the neutral and congruent conditions. It is evident that the tendency to substitute a different but logical word for the misidentified target (random completions) is higher in the neutral than in the incongruent condition and even higher in the congruent condition. On the other hand, the tendency to say "I don't know" (no response) is lower in the congruent and neutral conditions than in the incongruent condition. This tendency can be explained assuming that syntactic incongruence inhibited identification and enhanced uncertainty. The absence of syntactic incongruence in the congruent condition eliminated inhibition, and reduced uncertainty even when targets were misidentified. As a result, the percentage of random completions in the congruent condition was twice as large as the percentage of no responses.

The present results diverge from those reported by West and Stanovich (1986) who, using a similar neutral condition, found only inhibition. However, in addition to differences in task (West and Stanovich used a visual lexical decision task), the two studies differ in several other meaningful ways and, therefore, cannot be straightforwardly compared. First, we presented auditory masked-words whereas West and Stanovich (1986) used visually presented unobstructed stimuli. Although we have no evidence for a differential effect of

context in speech perception and reading, we cannot ignore this possibility. Moreover empirical findings on associative and semantic priming in reading, suggest that context effects are larger for degraded than for undegraded words (Stanovich & West, 1983). It is also possible that the divergence between the two studies is partly accounted for by differences between the material used in the two studies. In contrast to the semantically anomalous sentences used by West & Stanovich (1986), our sentences were always semantically sound.

Since we have no direct evidence about the influence of the above- mentioned factors on context effects and how they interact with syntactic priming, our ability to draw general conclusions is limited. Therefore inferences regarding the existence of facilitatory and inhibitory components to syntactic priming, and especially the finding of the equal contribution of the two components, may be restricted to the specific condition of the present demonstration. Despite this limitation, we can continue our general course and investigate the involvement of attention mechanisms with each of these two components.

EXPERIMENT 2

In the present experiment we examined the influence of presenting congruent and incongruent sentences in separate or mixed blocks on the inhibitory and facilitatory components of the syntactic priming effect.

Studies of semantic priming in visual word perception generally showed that lowering the proportion of related targets in the list reduced the amount of inhibition (Fischler & Bloom, 1979; Stanovich & West, 1981; but see Stanovich & West, 1983, Experiment 4). Within the framework of the two-process theory of Posner & Snyder (1975), most authors have assumed that the influence of the ratio between related and unrelated targets is mediated by attention mechanisms (e.g., Fischler & Bloom, 1985; Stanovich & West, 1983; Tweedy, Lapinski, & Schvaneveldt, 1977). Specifically it has been assumed that lowering the proportion of related targets discourages word perception strategies based on context-related expectations.

A similar manipulation was used to compare semantic vs. syntactic priming effects in visual word perception (Goodman et al., 1981 and Seidenberg et al., 1984). These studies suggested that the syntactic priming effect is mediated primarily by post-lexical, strategic mechanisms. In these studies, however, no attempt was made to examine the effect of separately manipulating strategies design to operate selectively on the facilitatory and inhibitory components of the syntactic priming effect. We applied the blocked vs. mixed presentation technique to disentangle the effect of attention mechanisms on each of these two components.

The blocked condition is an extreme case of manipulating the ratio between incongruent and congruent sentences, where the proportion of incongruent stimuli is either 1:0 or 0:1. This proportion was contrasted with a 1:1 ratio of incongruent and congruent stimuli used in the mixed condition. Therefore, the comparison between the blocked and mixed modes of presentation should maximize the effect of attentional processes that may mediate syntactic priming. A differential effect of the presentation mode on the percentage of correctly identified words in congruent and incongruent sentences should suggest that attention mechanisms are differentially involved in the mediation of the facilitatory and inhibitory components of the syntactic priming effect. Particularly, the involvement of attention mechanisms should reduce interference in the blocked presentation, leading to a higher percentage of identification of incongruent targets. On the other hand, the absence of an interaction between the modes of presentation and the congruity of the sentence should indicate that attention mediates the two components to a similar extent.

Method

Subjects. The subjects were 60 undergraduate students who did not take part in the first experiment. They participated in this experiment for course credit or for payment. They were all native speakers of Hebrew, without any known hearing problems.

Test Materials. The sentences were those used in Experiment 1, with the exception of the neutral stimuli. Thus each stimuli list included 40 sentences, 20 congruent and 20 incongruent. In the "mixed" presentation the 40 sentences were randomized and presented in one block. In the "blocked" presentation congruent and incongruent sentences were clustered separately in two blocks of 20 sentences each. The sentences in each of the two blocks were randomized.

A target appeared only once in each list (with the exception of sentences of Type 4, see above). Across lists, each target appeared equally in the congruent and incongruent conditions.

Procedure. Different 30 subjects were tested with each presentation mode. Subjects were randomly assigned to one of the lists, so that each subject was exposed equally to syntactically congruous and incongruous sentences.

The mixed presentation followed the same experimental procedures as in Experiment 1. The experimental list was preceded by a mixed list of 12 practice sentences (6 congruent and 6 incongruent).

In the blocked presentation, 15 subjects began with the congruent block, and 15 with the incongruent block. Each block was preceded by 8 practice sentences in the respective congruity condition. No special instruction were given before the incongruous block, but the "peculiar" structure of the sentences was not denied in reply to queries raised by the subjects following practice with incongruous sentences (as was true for the mixed condition as well).

Results

The percentage of correct identification of targets was averaged for each subject and target in each congruity condition. Separate means were computed for each presentation group. The percentages of correct identification of syntactically congruent targets was almost identical in the blocked and the mixed presentation groups. In contrast, more incongruent targets were identified in blocked than in mixed presentation (Figure 2).

The statistical significance of the observed differences was tested by two-factor analyses for subjects (F1) and for stimuli (F2). The factors were Congruity condition (congruent, incongruent) and Mode of presentation (mixed, blocked). Both main effects were significant [$F1(1,58)=486.7$, $MSe=123$, $p<.0001$, $F2(1,59)=128.7$, $MSe=937.3$, $p<.0001$, and $F1(1,58)= 18.1$, $Mse=192$, $p<.0001$, $F2(1,59)=21.6$, $Mse=296$, $p<.0001$, for the Congruity and Mode of presentation effects, respectively]. The most interesting result, however, was the significant interaction between the two factors, revealing that the presenting incongruent and incongruent sentences in separate blocks improved the identification of incongruent targets, but had no effect on congruent targets [$F1(1,58)=25.6$, $MSe=123$ $p<.0001$, $F2(1,59)=21.9$, $MSe=256$, $p<.0001$].

Errors in Experiment 2 were categorized and analyzed using the same types as elaborated in Experiment 1 (Table 2).

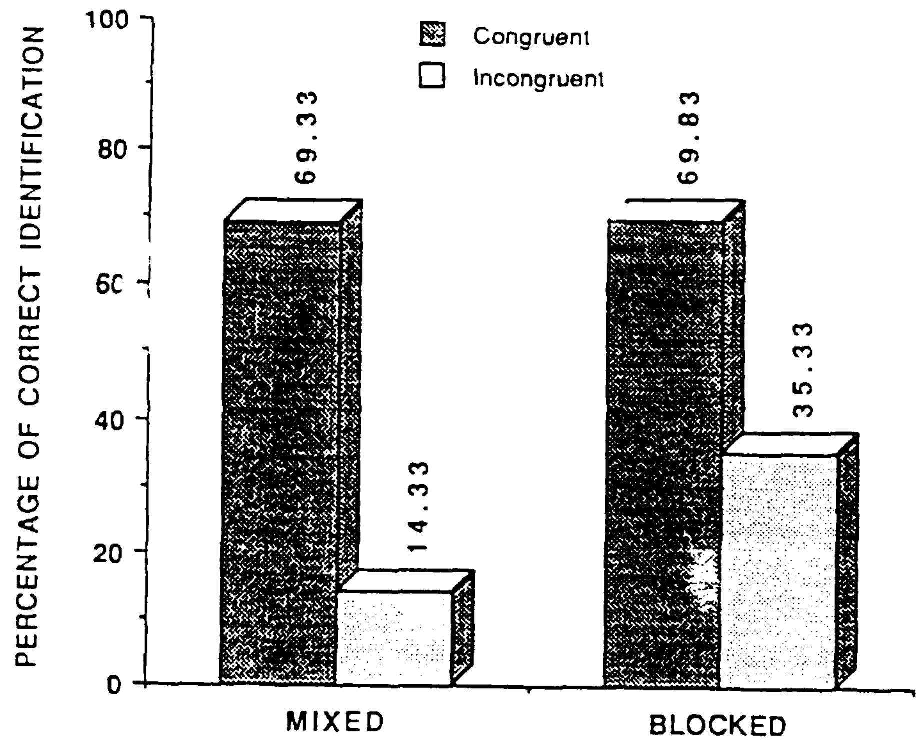

Figure 2. The percentage of correctly identified congruent and incongruent targets in the mixed and blocked presentation conditions.

Table 2. *Mean percentage of errors (SEm) of each type in each congruity condition in the mixed and blocked presentation modes.*

		ERROR TYPE			
CONGRUITY CONDITION		SELF CORRECTION	RANDOM COMPLETION	NONSENSE	NO RESPONSE
CONGRUENT	MIXED	–	63.7% (5.7)	0.4% (0.4)	33.9% (5.2)
	BLOCKED	–	56.6% (4.5)	2.9% (1.4)	37.1% (4.4)
	MIXED	17.5% (1.8)	22.2% (2.5)	4.0% (1.2)	56.4% (3.3)
INCONGRUENT	BLOCKED	13.9% (1.8)	30.2% (3.6)	11.8% (2.1)	43.4% (4.1)

In the congruent condition the distribution of errors was similar for mixed and blocked presentation modes (the interaction was not significant $F(2,116)<1.0$). Errors were unevenly distributed among types [$F(2,116)=70.9$ $MSe=729$, $p<.0001$]. The pattern of this distribution was similar to Experiment 1: There were significantly more random completion than no response errors ($p<.01$). In the incongruent condition, on the other hand, there was a significant interaction between the distribution of errors among the types and the mode of presentation [$F(3,174)=5.2$, $MSe=1532$, $p<.01$]. Post hoc analysis (Tukey-A) revealed that, although significantly less correction than no response errors were made in both presentation modes ($p<.01$), the difference was larger in the mixed than in the blocked presentation.

Discussion

The present results revealed that manipulating the proportion of congruent and incongruent sentences in the experimental list affects only the inhibitory component of the syntactic priming effect. In comparison to a mixed presentation (1:1 proportion), the presentation of incongruent and congruent sentences in separate blocks reduced the amount of inhibition without altering the amount of facilitation. Assuming that this manipulation influences primarily strategic components, the present results suggest that syntactic priming includes attention-mediated mechanisms that are reflected more in its inhibitory than its facilitatory effects.

An attention mechanism that might have been affected by our manipulation is the strategic process of generating context-based expectations about the target's syntactic form. The application of this strategy should probably be encouraged by a high proportion of congruent sentences in a mixed list and discouraged by frequent syntactic incongruence. Hence, the tendency to generate expectations (leading to less identification of incongruent targets) should decrease in parallel to the reduction of the percentage of congruent sentences in the list. Informal study of the percentage of correctly identified incongruent targets across Experiments 1 and 2, conformed to this prediction: Incongruent targets were identified least (14.3%) in the mixed condition of Experiment 2 where 50% of the sentences were congruent, more in Experiment 1 (27.3%), where, due to the neutral condition only 33% of the sentences were congruent, and most in the blocked condition of Experiment 2 (35.3), where there were no congruent sentences. In contrast, the proportion of congruent sentences did not affect the percentage of correctly identified congruous words significantly (69.3%, 74.8%, 69.8% in the mixed presentation of Experiment 2, Experiment 1, and the congruent block of Experiment 2, respectively) This suggests that the facilitatory component of the syntactic priming effect is less sensitive to strategic mediated processes.

Additional support to our interpretation is provided by the distribution of errors among the

different types. A comparison between the mixed and blocked presentation modes revealed that the percentage of random and nonsense errors (those that reflected less concern about the priming sentence) was higher in the blocked than in the mixed presentation modes, whereas the opposite trend was observed for no response and self correction errors (that reflect the influence of the priming effect induced by the syntactic structure of the sentence). Hence, it appears the syntactic context influence on word identification was reduced in the blocked relatively to the mixed presentation mode. The singularity of this interaction to the incongruent condition is in agreement to our hypothesis that the generation of expectations is one of the factors involved in producing the syntactic priming effect on word identification.

It is worth noting that the present results diverge from the results of Stanovich & West (1983), who found that the pattern of contextual (semantic) effects was not altered by increasing the proportion of congruent targets. This divergence may either suggest a fundamental difference between the involvement of attention in semantic and syntactic context effects, or that our manipulation of blocking congruity condition was more powerful than changing proportion of congruent and incongruent targets within a mixed block.

In Experiment 3 we used a different method to manipulate the subjects' tendency to generate expectations as a strategy of word identification, in an attempt to corroborate the differential involvement of attention with the facilitatory and inhibitory components of the syntactic priming effect.

EXPERIMENT 3

In contrast to Experiment 2, where our manipulation was meant to discourage the generation of expectations for specific syntactic forms, in the present experiment we sought to encourage this strategy.

Studies of semantic priming revealed that the length of the inter-stimulus interval (ISI) [or the stimulus onset asynchrony (SOA)] between the context and the target, influences the relative weight of the attention-based component of the priming effect with single-word (Antos, 1979; Neely, 1977) and sentence contexts (Stanovich & West, 1979). Different ISIs were used in different studies and the general consensus among authors is that, within a limited range of times, the tendency to use context-based expectations increases with longer ISIs. Possibly, at longer ISIs the subject has more time to process the context and generate such expectations.

The influence of the ISI between context and target on syntactic context effects is not as clear. For example, using a lexical decision task with printed Serbo-Croatian stimulus-pairs, Lukatela et al., (1982) found significantly larger syntactic priming effects when the SOA was 800 ms than when it was 300 ms. However, with auditory presented stimuli (in Serbo-Croatian), Katz et al. (1987) did not find a reliable interaction between the length of the ISI (0 vs. 800 ms) and the magnitude of the syntactic priming on lexical decision. Despite the apparent divergent results, both groups of authors suggested that the syntactic context effect reflects the operation of an autonomous automatic module rather than an attention mediated mechanism. However, as Katz et al. (1987) pointed out, it is possible that this conclusion holds only for the particular case of inflectional morphology characteristic to Serbo-Croatian. Indeed, indirect evidence for non-automatic aspects of syntactic priming has been found in English (Tanenhaus, et al., 1979). Using a naming task, these authors reported that at 0 ms SOA, subjects were insensitive to the specific syntactic (and semantic) form of the prime, whereas at 200 ms, the targets were facilitated only by appropriate forms. Concluding these results Tanenhaus et al. (1979) suggested that at longer SOAs, syntactically inappropriate forms are inhibited by veiled controlled process (i.e., Shiffrin & Schneider, 1977). The time course of the controlled process, however, was obscured by the finding that at 600 ms SOA, it's effect was not as evident as at 200 ms SOA. Together, the previous studies cannot unequivocally support or reject the existence of attention-mediated components of the syntactic priming effect. An additional step towards the clarification of the role of attention in syntactic priming can be made by distinguishing between effect of ISI manipulation on the inhibitory and facilitatory components of syntactic priming

In the present experiment we used two ISIs. One was set at the normal speech rate, and the other was 350 ms.[2] On the basis of the results of Experiment 2, we anticipated that the ISI manipulation should affect primarily the inhibitory component. More specifically we predicted that the at the longer ISI, syntactic incongruity should have a more deleterious effect on the identification of targets than at normal speech rate whereas the facilitatory effect will not change.

Method

Subjects. Sixty subjects participated in this experiment. Thirty were the mixed presentation group from Experiment 2. The other 30 were naive undergraduates who did not take part in the previous experiments, and participated in this experiment for course credit or for payment. All the subjects were native speakers of Hebrew, without any known hearing problems.

Stimuli and Design. The stimuli lists were those used in the mixed presentation condition of Experiment 2. The only alteration was the introduction of a silence period of 350 ms between the offset of the last unmasked word in the context and the onset of the masked target. The 30 naive subjects were tested with these lists. Their performance was compared to the performance of the mixed presentation group in Experiment 2, who heard the same lists at a normal speech rate. Each subject was exposed equally to syntactically congruous and incongruous sentences. Thus, the subject analysis was a mixed model ANOVA. The ISI effect was tested between groups and the syntactic congruity effect within subjects.

Across subjects, each target appeared equally in the congruent or incongruent conditions, and at each ISI. Thus the stimulus analysis was completely within stimulus.

Procedure. The experimental procedure of the present experiment in which we tested only the longer ISI condition was the same as that followed in the mixed presentation condition of Experiment 2. The test list was preceded by 12 practice sentences that included the silence interval. Except of being informed about the brief silence period preceding the masked target word, the subjects were instructed identically as in the mixed presentation condition of Experiment 2.

Results

The percentage of correct identification of targets was averaged for each subject and each target in each congruity condition. These results were compared to the percentage of correct identification of congruent and incongruent targets at normal speech rate in the mixed presentation condition of Experiment 2 (Figure 3). Congruent targets were identified almost identically in the two ISI conditions. In contrast, the percentage of incongruent targets identification was smaller in the 350 ms ISI condition than at normal speech rate.

The statistical significance of the observed differences was tested by two-factor analyses (mixed model for subjects (F1) and repeated measures for stimuli (F2)). Both the congruity and ISI main effects were reliable [$F1(1,58)=848.1$, $Mse=123$, $p<.0001$, $F2(1,59)=232.7$, $Mse=880$, $p<.0001$, for the congruity effect and $F1(1,58)= 5.2$, $Mse=159$. $P<.0264$, $F2(1,59)=7.412$, $Mse=268$, $p<.0085$ for the ISI effect]. The most important result, however, was the reliable interaction between the two factors, revealing that the 350 ms silence interval reduced the identification of incongruent targets, but had no effect on congruent targets [$F1(1,58)=4.1$, $Mse=123$ $p<.0488$, $F2(1,59)=4.4$, $Mse=211$, $p<.0411$].

The distribution of errors in the different ISI conditions is presented in Table 3.

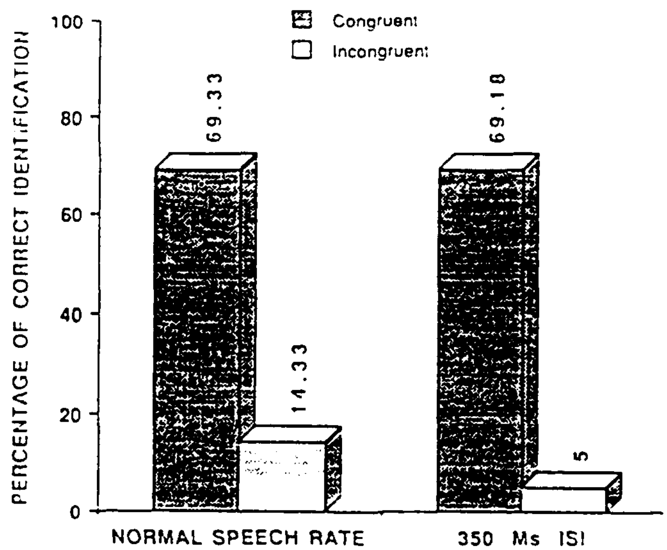

Figure 3. **The percentage of correctly identified congruent and incongruent targets at normal speech rate and with 350 ms ISI between context and target.**

Table 3. *Mean percentage of errors (SEm) of each type in each congruity condition with normal speech rate and with 350 ms ISI between context and target.*

		ERROR TYPE							
CONGRUITY CONDITION		SELF CORRECTION		RANDOM COMPLETION		NONSENSE		NO RESPONSE	
CONGRUENT	NORMAL	–		63.7%	(5.7)	0.4%	(0.4)	33.9%	(5.2)
	350 ms	–		53.1%	(5.2)	1.8%	(1.2)	45.1%	(4.8)
INCONGRUENT	NORMAL	17.5%	(1.8)	22.2%	(2.5)	4.0%	(1.2)	56.4%	(3.3)
	350 ms	15.2%	(1.5)	17.9%	(2.5)	1.9%	(0.9)	65.4%	(3.2)

The ISI manipulation influenced the distribution of errors in the incongruent condition [$F(3,174)=2.72$, $MSe=209$, $p<.05$], but not in the congruent condition [$F(2,116)=2.15$ $MSe=833$, $p>.12$]. Across conditions the distribution of errors was similar to that observed in the former two experiments and significant [$F(3,174)=179.3$, $MSe=209$, $p<0.0001$ and $F(2,116)=61.2$, $MSe=833$, $p<.0001$, in the incongruent and congruent conditions, respectively]. Post hoc analysis (Tukey-A) of the interaction revealed that, while no response type errors were more abundant in the 350 ms ISI condition than with normal speech rate, the percentage of all other three error types was reduced in the latter than in the former condition.

Discussion

Increasing the ISI from a normal speech rate to 350 ms between the context phrases and the targets, reduced the percentage of correct identification of incongruent targets but had no influence on the identification of congruent targets. These results confirmed our previous observations that the facilitatory and inhibitory components of the syntactic priming effect are differentially sensitive to the manipulation of attention-based strategies of word identification.

In Experiments 3, as well as in Experiment 2, our manipulation affected only the inhibitory priming component albeit, in each experiment in an opposite direction. Therefore, these results suggest that in both experiments we manipulated the same attention-mediated priming process. Assuming that this process involves the generation of context-based expectations, the results of both experiments support our distinction between an inhibitory component of syntactic priming, which reflects an attentional process of generating expectations, and a facilitatory component, that is less reliant on attentional mediation.

The distribution of errors is in complete agreement with the above interpretation. Again, the ISI manipulation influenced the distribution of errors only in the incongruent condition. However, the trend of this interaction was opposite to that found in Experiment 2. Whereas discouraging the generation of context-based expectations in the blocked, relatively to the mixed presentation mode increased the percentage of random and nonsense error-types, encouraging such a strategy by introducing a longer ISI lead to a decrease of such errors while increasing the percentage of no response errors.

Despite the correspondence between the results of the two experiments and the coherence of the emerging picture, the ISI manipulation should be considered with caution. Previous studies of the time course of sentence-context effects on word perception are not conclusive. For example, Fischler and Bloom (1979; 1980) presented written sentences word by word, manipulating the presentation rate. Contrary to our results, they found almost no facilitation of lexical decision for expected target words while the inhibition of incongruous targets was evident at all presentation rates. Their conclusion was that the effect of the sentence semantic context on word recognition is limited to an inhibitory postlexical process. This inhibition is probably related to the sentences' semantic incongruity and is not sensitive to the manipulation of ISI. A closer look at their data, however, reveals that, in agreement to our results, the magnitude of the inhibition effect on lexical decision (speed and accuracy), was twice as large at the slower rates (4 and 12 words per second), than at the higher presentation rates (20 and 28 words per second).

One problem in analyzing the ISI effect is that different studies manipulated different time intervals. It is possible that the ISI influence is not monotonic, and that it differs with factors such as task, presentation modality, and the linguistic context that it is investigated. It is, therefore, possible that relatively small differences in the particular ISIs compared in different studies, account for the variation of the results. The results of two pilot experiments that preceded the present study support this possibility. In these pilot experiments we explored the effect of 500 and 1000 ms ISI compared to normal speech rate. The effect of syntactic priming at these ISIs were not reliably different than at normal speech rate. An interesting trend emerged however across the ISIs. At 1000 ms, the increase in the inhibition was accompanied by a decrease in the facilitation. Relative to 1000 ms, 500 ms ISI caused a smaller decrease on the magnitude of the facilitation and an even bigger increase in the inhibition effect. Finally as reported in the present experiment, 350 ms ISI had no effect on the magnitude of the facilitatory component while significantly increasing the magnitude of the inhibition. Thus, it appears that the interaction between the ISI and the syntactic context effect is limited to a specific range. This limit might also account for the absence of a difference between the syntactic congruity effect at 0 and at 800 ms ISI in Katz et al., (1987) study. Despite the caution, however, the present results suggest that ISI manipulation, when carefully applied, may reveal interesting aspects of the context effects.

The inherent problems of ISI manipulation are not essential, however, to our conclusions regarding the involvement of attention in mediating syntactic priming effects. Therefore, we may resume our discussion of the relation between attention mechanisms and the syntactic context effects as revealed in the present study.

GENERAL DISCUSSION

In the present study we examined the inhibitory and the facilitatory aspects of syntactic priming as it is reflected in the identification of auditory masked targets that were presented as last words in clearly displayed sentences. In Experiment 1, we found evidence for both components. In addition, the data indicated that, at least for the present experimental conditions, facilitation and inhibition contribute equally to the syntactic priming effect. In Experiments 2 and 3 we found that manipulation of attention-related factors affected the magnitude of the inhibition but had no effect on facilitation. The presentation of congruent and incongruent sentences in separate blocks attenuated inhibition relative to a mixed condition. On the other hand, Experiment 3 suggests that the insertion of 350 ms of silence between the target and the context amplified the inhibition relative to normal speech rate.

Across experiments, the scarcity of self-corrections and the abundance of no-response errors relatively to random completions in the incongruent condition, on one hand, and the increased percentage of random completion errors at the expense of no-response errors in the congruent condition on the other hand, discarded the possibility that the variation in the percentage of correct identification between the different congruency conditions simply reflected a strategy of intelligent guessing on the basis of partially identified information. Taken together our results indicate that the syntactic context effects observed in the present study were probably related to a post-lexical syntactic analysis of the input, whose possible nature is discussed bellow.

In accord with the commonly-held account for attention-mediated factors in semantic priming (Fischler, 1977; Fischler & Bloom, 1979; Neely, 1977; Stanovich & West, 1981, 1983), we suggest that our manipulations influenced an attention-based mechanism that mediates the generation of expectations. However, the concept of generating expectations in the syntactic domain can not simply be an extension of the models suggested to account for attention mediation in semantic priming.

When analyzing discourse the subject naturally expects that the input will be coherent with his/her existent linguistic knowledge (deGroot, Thomassen & Hudson, 1982; Fischler & Bloom, 1980). We assume that this strategy is applied in the syntactic as well as in the semantic domain. Specifically, we assume that when a particular syntactic structure is alluded by the context, the perceiver expects grammatical forms that can be integrated into that structure. The observed inhibition of incongruent targets in the present study might have been caused by the violation of those expectations. Possibly, incompatible input induces a second pass analysis of the target and/or context. This additional process may delay or, when the target is degraded, may suppress its identification.

In contrast to previous studies of syntactic priming (e.g., West & Stanovich, 1986), we found that the identification of congruent targets was

facilitated relative to a neutral condition. Regardless of the particular reasons for this discrepancy (that have been discussed in Experiment 1), we may speculate on possible sources of this facilitation. In the semantic priming domain facilitation is assumed to reflect two different processes: One is an automatic spreading of activation among nodes related in the semantic network (Collins & Loftus, 1975). The second is the confirmation of an explicit prediction regarding target identity (Becker, 1980). Because the existence of a syntactical organized network is supported neither by empirical evidence nor by theoretical considerations, the mechanism of spreading activation is an improbable source of facilita ion in syntactic priming. Therefore the facilitation of syntactically congruent targets is better explained by a process that relates on ad hoc generated syntactic structures. The mechanism of explicit prediction suggested by Becker (1980) cannot be directly applied to syntactic priming because, in our study, the identity of the target could not be predicted by the context (for similar claims see also Oden & Spira, 1983; Tanenhaus et al., 1979; Tyler & Wessels, 1983). Therefore, we are lead to believe that the mechanism of facilitation by syntactic priming is based on the same form of expectations as postulated to account for the inhibition of incongruent targets. Thus, we propose that the same expectations may be used by different mechanisms to exert both facilitation and inhibition on the identification of the target. The first, which was postulated above, causes the inhibition of incongruent targets. The second, may facilitate the identification of congruent targets either because expected structures may assist the integration of the sentence, or because they reduce the amount of sensory input needed for the identification of a word.[3]

The above proposal that, in syntactic priming, both the facilitation and the inhibition are based on a similar process of generating expectations, apparently contradicts with the results of Experiments 2 and 3 that showed that manipulating the tendency to generate expectations affects only the inhibition. This contradiction can be resolved assuming that the generation of expectations at the sentence level is motivated by the natural assumption of syntactic coherence (similar claims related to the processing of sentences at the semantic level were made deGroot (1982) and by Fischler & Bloom, 1980). It is conceivable that the tendency to generate expectations is not under strategic control. This view is compatible with the residual inhibition observed in the incongruent block, which suggests that despite the clear incongruent structure of all sentences, the initial expectations could not be completely avoided. Hence, at the sentence level, the expectations are probably generated by a veiled controlled process which uses only minimal attention resources (Schneider & Shiffrin, 1977). Such a process probably stands at the basis of the facilitatory mechanism of syntactic priming. On the other hand, as discussed above, when the same expectations are violated by incoherent input, attention is mobilized to trigger and control the additional, post-lexical process of re-evaluation, which we suggest that it is the main mechanism of the inhibition. Consequently, strategic changes should influence the magnitude of the inhibition, but have only minimal effect on the facilitation. Indeed, the interaction between the distribution of errors and the presentation procedure found in Experiments 2 and 3 only in the incongruent condition supports this view. Attenuating the tendency for re-evaluation of context-based expectations (in Experiment 2) reduced no-response and self-correction errors and increased the percentage of random and nonsense responses. On the other hand, facilitating the generation of context-based expectations (in Experiment 3) increased subjects' uncertainty as manifested by the increase in the no-response type errors. Should this process influence lexical access rather than post-lexical re-evaluation, the opposite manipulations on subjects' strategies in Experiments 2 and 3 should have had an effect on the overall percentage of correct identification in the incongruent condition but not on the distribution of errors. Our hypothesis that a similar attention mechanism is the basis of both the facilitation and the inhibition of performance in the syntactic priming task, also implies that the allocation of attention, at least in language processing is not an all-or-none phenomenon. Rather, based on data-driven or pre-determined strategies different amounts of attentional resources are directed to the different aspects of language perception processes.

A caveat of the above discussion is that the task used in the present study required the identification of degraded stimuli. Therefore is possible that the magnitude of the syntactic priming was largely dependent upon these conditions. In particular, the inhibition might have been much smaller if the auditory input was clear. The need for the re-evaluation could have been less conspicuous in the absence of auditory

uncertainty. However, we believe that using degraded stimuli we were able to tap mechanisms of top-down processing of syntax that are available to the language speaker.

REFERENCES

Antos, S. J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 527-545.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*, 493-512.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 389-401.

Bentin, S., Deutsch, A., & Liberman, I. Y. (1990). Syntactic competence and reading ability in children. *Journal of Experimental Child Psychology, 48*, 147-172.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition, 16 (3)*, 187-195.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.

deGroot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. W. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10 (4)*, 358-370.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition, 5*, 335-339.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 1-20.

Fischler, I., & Bloom, P. A. (1980). Rapid processing of the meaning of sentence. *Memory & Cognition, 13 (2)*, 128-139.

Fischler, I. R., & Bloom, P. A. (1985). Effects of constraint and validity of sentence contexts on lexical decisions. *Memory & Cognition, 13 (2)*, 128-139.

Fodor, J. A. (1983). *The modularity of mind.* MIT Press.

Goodman, G. O., McClelland, J. L., & Gibbs, JR., R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition, 9 (6)*, 580-586.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T.(1985). Grammatical priming in inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (1987). Grammatical information effects in auditory word recognition. *Cognition, 25*, 235-263.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming in inflected nouns. *Memory & Cognition, 11(1)*, 59-63.

Lukatela, G., Moraco, J., Stojonov, D., Savić, M. D., Katz, L. & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research, 44*, 297-311.

Marslen-Wilson, W. D. (1980). Speech recognition as a psychological process. In J. C. Simon (Ed.), *Spoken language generation and recognition*. Dordrecht: Reidel.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception & Performance, 7*, 634-644.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. Rabbitt & S. Dornic (Eds.), *Attention and Performance, Vol. 5*, New York: Academic Press.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition, 4*, 648-654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106 (3)*, 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition : A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*, Hillsdale, NJ: Erlbaum.

Oden G. C. & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology, 35A*, 51-64.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. Rabbitt & S. Dornic (Eds.), *Attention and Performance, Vol. 5*, New York: Academic Press.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and general theory. *Psychological Review, 84*, 127-190.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P., (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12 (4)*, 315-328.

Stanovich K. E., & West R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition, 7 (2)*, 77-85.

Stanovich K. E., & West R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7 (3)*, 658-672.

Stanovich K. E., & West R. F. (1983). On priming by sentence context. *Journal of Experimental Psychology: General, 112 (1)*, 1-36.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in processing ambiguous words in syntactic context. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440.

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in appropriate context. *Memory & Cognition, 5*, 84-89.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34(5)*, 409-420.

West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition, 14 (2)*, 104-112.

Wright, B. & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12 (1)*, 31-45.

FOOTNOTES

Cognition, in press.

† Department of Psychology and School of Education, The Hebrew University, Jerusalem.

[1] Take, for example the sentence "A nice boy eats" which translated into Hebrew would sound "Yeled yafeh ochel". The morphological unit "yeled" (boy) contains information about gender (masculine) and number (singular). The same root with different affixes is used to form the word "Yaldah" (girl) or change the number. The agreement rule requires that the attributes and predicate will agree with the subject in gender and number: "yafeh" (nice) is a singular masculine form as is "ochel" (eats). The sentence "Yaldah yafah ochel" is a possible syntactic violation of that sentence because the predicate is in

masculine form while both the subject and attribute are in feminine form.

[2]This particular ISI was chosen on the basis of pilot studies. In the present study we were concerned to demonstrate the ISI effect on the two components of the syntactic priming effect and not to examine its precise time course of the putative controlled component. Therefore we examined different ISIs (1000 ms, 500 ms, and 350 ms), but completely analyzed only the later that had the most conspicuous effect.

[3]A similar model was proposed within the frame of the cohort theory (Marslen-Wilson, 1980). According to this model the syntactic context may facilitate word identification by limiting the size of an initial cohort to those members which belong to a single form-class category (Tyler & Wessels, 1983).

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 221-224

Starting on the Right Foot*
A review of Marilyn Jager Adams' *Beginning to Read: Thinking and Learning about Print***

Donald Shankweiler†

Marilyn Jager Adams has performed a valuable service to all who wish to improve how reading is taught. Her book presents a comprehensive and scientifically responsible treatment of problems of immense social importance—problems that partly because of their very complexity are too often treated cavalierly. This book is required reading for professionals engaged in research on design and assessment of programs of reading instruction and research on diagnosis and treatment of reading disability. It is also a valuable resource for a wider readership in psychology, cognitive science and education. Indeed, anyone who needs a clear-headed synthesis of relevant research findings bearing on the problems of learning and teaching to read can profit greatly from this book. With unusual thoroughness, Adams has reviewed the mass of research literature that bears on the debate between advocates and adversaries of the code emphasis in reading instruction. The tone is always constructive. She avoids the rancor that so often accompanies discussion of these issues. Though even-handed in her treatment, Adams does not wrap herself in the cloak of the eclectic; after sifting the evidence, she draws strong conclusions and states them boldly.

This book originated with a mandate from the United States Congress for a new appraisal of the place of phonics in teaching children to read. Inundated with complaints about the performance of the schools in imparting literacy, and confused by the welter of conflicting voices from the experts, Congress enacted legislation that led ultimately to the U.S. Department of Education's commission of this report. Responsibility for producing the report was placed in the hands of the Center for the Study of Reading, University of Illinois at Urbana-Champaign. Adams, a cognitive and developmental psychologist at the Center's branch at Bolt, Beranek and Newman in Cambridge, Massachusetts, was chosen for the task.

Given Adams' extensive background in investigation of basic reading processes, she was a logical choice and the choice proves to have been an excellent one. Charged with the responsibility for presenting a thoroughgoing clarification of the issues that divide the two sides in what Jeanne Chall has called "the great debate," Adams was given a free hand to shape the report. A panel consisting of well-known reading experts from around the nation was assembled to offer advice and criticism of interim drafts, but the book was written by Adams, not the committee. And to her great credit, the book is highly readable. It has none of the dryness one often finds in a technical report. The book displays a graceful and informal writing style and betokens an uncommon ability to use the language well.

As Adams points out, this book has a predecessor: the task of reviewing the relevant research literature was undertaken in the 1960s by Jeanne Chall whose report was published nearly 25 years ago (Chall, 1967). Appropriately, Adams often refers to the earlier work. It, too, was a praiseworthy review, but time does not stand still. The unprecedented technological explosion in the work place presents ever greater demands on reading skills. Moreover, the crisis in the schools has intensified, consensus on a remedy for the unacceptably high rate of illiteracy in our society seems as elusive as ever.

In the meantime, research activity has mushroomed both in quantity and in variety. An important new development since Chall's book appeared is the rediscovery of reading as a central

problem for investigation by mainline psychology. No less significant, reading and orthography have become major concerns within the fast-growing fields of applied linguistics and the psychology of language. One consequence of the remarkable surge in research on reading is obvious: Anyone who would undertake to review the literature must be prepared to digest and critically evaluate an enormous range of material. Accordingly, heavy demands are placed on a reviewer's knowledge and critical judgment. On the whole, Adams proves more than equal to the task.

The report has five parts. Part I deals with the nature of writing systems, the origin of the alphabet and the place of word recognition in reading. Part 2 presents the rationale for approaches to instruction that emphasize phonics, and it reviews research that attempts to compare the efficacy of this approach with other approaches. Part 3 presents conceptions of reading from the standpoint of laboratory analysis of what skilled readers do. It presents a model of the reading process that encompasses each of the components of reading skill and their integration in the act of reading. Part 4 articulates the goals of instruction in reading from the standpoint of the analysis of the skills of the mature reader presented in Part 3. Part 5 discusses research on the processes involved in learning to read. Part 6 summarizes the conclusions reached from the review of the research literature and discusses the implications for teaching and learning to read.

Adams begins with a discussion of the nature of writing. It is noted that true orthographies, unlike picture writing, represent words, and not meanings directly. This is an appropriate starting point because it underscores the key significance of the word in reading. The importance of apprehending each and every word in the text cannot be taken for granted, because it is unfortunately true that some popular programs of beginning reading instruction encourage the novice to skip words or to guess in the search for meaning. Adams leaves us in no doubt where she stands: This is bad advice for a beginning reader or anyone else. "Unless the processes involved in individual word recognition operate properly, nothing else in the system can either (p. 3)." The ability to identify printed words is necessary but not sufficient for reading; it must be backed up by well-oiled mechanisms of language comprehension. Reading depends on a system of skills whose components must mesh properly.

Alphabetic forms of writing are codes on the phonological structure of the language, or more properly, the morphophonological structure. By using letters to represent the several dozen consonant and vowel sounds of the language, alphabets achieve their great advantages over other forms of writing: First, economy—a small set of symbols is sufficient to represent any and all words in the language; second, transparency—a user who knows how the system works can usually recognize words in print that were previously known only through spoken language. Adams' account notes that these advantages come at a cost that must be borne by the beginner. Every alphabetic system presents its users with a problem of cognitive penetrability. Because vowels and consonants are co-produced and overlapped in time, these abstract phonemic units are not realized in speech as physically separable chunks of sound. That is probably one reason why they are often difficult to apprehend consciously (Liberman, Shankweiler, Fischer, & Carter, 1974). For the purposes of speaking and listening, language users need not attain awareness of phonemes. But to grasp the principle (by which alphabetic writing represents the phonemes and morphophonemes of the language), a would-be reader must first identify the speech units that the letters represent. Consequently, the grasp of the alphabetic principle is a rather sophisticated intellectual achievement.

Because the orthography of English is complex and often irregular, some commentators have overlooked that it is, nonetheless, essentially alphabetic. Adams does not make that mistake. Yet to dwell on the irregularities, as she does at the end of Chapter 2, is to invite a reader who is less than astute to draw the wrong conclusion and to miss the larger point: that there is a system to be learned and that, even in English, knowledge of the orthography is productive.

The chapters that follow present a much needed and thoughtful analysis of the pertinent information on phonics and reading. As for phonics, the term itself has long been a source of confusion. For the most part, Adams uses the term simply to denote instruction aimed at instilling the alphabetic principle. Well and good. But unfortunately the term has other connotations that are hard to shake off: In the minds of some people, phonics denotes an old-fashioned and discredited method of teaching reading by having children attempt to recognize a word by speaking the "sound" of each letter. The method implies that what a reader does is to approach words piecemeal by translating the letters that make up a word into their phonetic equivalents, letter by

letter, as though reading were simply spelling aloud. Thus the term phonics has come to represent an inapt caricature of the reading process. Accordingly, Liberman and Liberman (1990) recommend substituting for *phonics* Chall's term, *code-based approach*.

As Isabelle Liberman (who is cited by Adams on this point) often explained, letter-by-letter encoding is assuredly not what a successful reader does. The word *bat* contains one syllable, not three; the word is not *buh-a-tuh* but *bat*. Yet some beginning readers will say something like "*buh-a-tuh*" when asked to read the word and will never manage to discover that the word is *bat* (Liberman, 1973). In Adams words, "It is as though these children can find no connection between the sequence of sounds they have produced and the highly familiar word which they have 'read.' It is not enough to have memorized the sounds that go with each letter. To make use of those sounds, the child must realize that they are the subsounds of language" (p. 208). Beginners who are stuck in this way can be helped to develop phonological awareness, that is, to become aware of the phonological structure of words, by identifying their phoneme and syllable constituents. Then they are prepared to grasp the alphabetic principle and can begin to build word recognition skills on a solid foundation. As Adams notes, experienced readers parse the letter strings, ordinarily apprehending sequences of letters that correspond to a demi-syllable at minimum. According to laboratory research discussed in Part 3, such sequences constitute the major spelling patterns that experienced readers implicitly recognize as wholes.

Spelling patterns must be not only apprehended but also overlearned to the point that word recognition can become unhesitating and automatic. Speed, as well as accuracy, is important because the fast-fading short-term memory forms the stage for the integration of words into syntactic units. If word decoding routines work poorly, all other aspects of reading will be hampered and comprehension will be correspondingly poor, a point often stressed by Perfetti and his associates (Perfetti, 1985). Thus, although word recognition *per se* is not the goal of reading, getting the meaning of the text depends on it. And word recognition, in turn, depends on accurate identification of the lower-level building blocks: the letters and the spelling patterns formed by letter combinations.

In Part 3, Adams sketches a model of reading that derives largely from the work of Seidenberg and McClelland. The chief characteristic of this model is that information the reader derives from print interacts freely and at every level with stored knowledge. Thus the model contrasts with a hierarchical model in which information flow is largely unidirectional and bottom-up. Other researchers have maintained that an interactive model does not readily account for the important differences between reading and speech perception. Above all, it offers no explanation of the fundamental fact that speech is acquired by every neurologically normal child whereas reading skill is far from universally acquired. For some researchers, a unidirectional model seems dictated by the modular nature of the language apparatus (see Crain, 1989; Fodor, 1983; Shankweiler & Crain, 1986). Of course the question is not whether linguistic input (whether speech or print) must make contact with stored knowledge, but how and when. The modular view supposes that processing within the language module is accomplished before the linguistic input is integrated with other aspects of cognition. On this account, it is emphasized that word recognition by ear is privileged in the sense that it is served by mechanisms that evolved in our species and that form part of a coherent biological specialization for language. In contrast to speech, the alphabet is an artifact. Learning to use it is a cognitive task in a way that primary language acquisition is not. It has been argued that an adequate theory of reading would have to explain the difficulty of reading and the comparative ease of acquiring a spoken language (Liberman, 1989).

After examining the myriad studies comparing programs for the teaching of beginning reading, Adams concludes that the great majority of program comparison studies indicate that approaches that incorporate code-based instruction "...result in comprehension skills that are at least comparable to, and word recognition and spelling skills that are significantly better than, those that do not" (p. 49). This, she notes, is exactly the same conclusion that Jeanne Chall drew 25 years earlier. Code-based approaches that help the beginner to appreciate that words have an internal phonological structure and to recognize that word spellings represent that structure have the edge over programs that pass over these aspects.

While stressing that these program comparisons are essential, and have been highly informative, Adams is sensitive to the limitations of these research studies and in Chapter 3 she knowledgeably discusses the reasons why they so

often yield noisy data. The classroom teacher, who is charged with implementing the program, is often the weak link. Adams' conviction that successful readers must grasp the alphabetic principle and that code-based teaching is the best way to help beginners to grasp it stems only in part from such program comparisons. At least as important are other research findings which are discussed in detail in this book. The pertinent evidence comes from a variety of sources: It includes the findings of research on prereaders, prediction studies seeking to identify those preschoolers who are at risk for reading failure, follow-up studies on the long-term educational consequences of failing to crack the code in the early primary grades, studies identifying the shared characteristics of unsuccessful readers, and finally, the picture of reading derived from research on the skilled reader. Adams concludes that all these lines of evidence converge in underscoring the vital importance of helping children grasp the alphabetic principle from the beginning. This entails giving prereaders adequate preparation for learning to read by instilling phonological awareness (introducing, through well-chosen word games, the fact that words have an internal phonological structure), and by demonstrating to beginning readers, through examples, how the spelling of a word represents its phonology.

Of course, some children will infer the principle with little guidance from anyone and will make rapid progress in word recognition skills. But for a significant minority, which includes some children from highly favorable home backgrounds as well as many from unfavorable home environments, extensive instruction is needed to compensate what appears to be a general weakness in the phonological component of language. Unfortunately, these are the very children who are often deemed unable to profit from such instruction and are therefore denied access to it.

If the case for code-based instruction is unassailable, why, then, is it so often resisted? Adams ponders this question near the end of the book. She is inclined to think that the reason is that it is often poorly implemented in practice. Implementation, she notes, depends on clarity with respect to goals; the teacher must understand why each activity is included. "It is with respect to principles and goals that I would most strongly fault the major reading curricula" (p. 423). Certainly, one cannot disagree that it is vitally important for teachers to understand what they are attempting to accomplish through their teaching, and that a recipe book or a manual, no matter how logically ordered and detailed, will not impart that knowle e. The problem will not be easy to solve. There much ignorance concerning the needs of beginning readers both on the part of teachers and teachers of teachers. Adams' book takes many constructive steps toward remediation of ignorance about reading. Let it be read and reflected upon in every place where teachers of reading are taught, and may it shine like a beacon!

REFERENCES

Chall, J. (1967). *Learning to read: The great debate*. New York: McGraw Hill.

Crain, S. (1989). Why poor readers misunderstand spoken sentences. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle*. (IARLD Monograph Series). Ann Arbor, MI: University of Michigan Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Liberman, A. M. (1989). Reading is hard just because listening is easy. In C. von Euler, I. Lundberg & G. Lennerstrand (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading* (pp. 197-205). London: Macmillan.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

Liberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Annals of Dyslexia, 40*, 51-76.

Liberman, I. Y., Shankweiler, D., Fischer, F. W. & Carter, B. (1974). Explicit phoneme and syllable segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Shankweiler, D. & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 24, 139-168.

FOOTNOTES

*Appears in *Educational Research, May*, 33-35 (1991).

**Cambridge, MA: Bradford Books/MIT Press (1990). 494 pp. ISBN 0-262-01112-3.

†The University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 225-250

Null Subject vs. Null Object: Some Evidence from the Acquisition of Chinese and English*

Qi Wang,† Diane Lillo-Martin,† Catherine T. Best,†† and Andrea Levitt†††

Since young English-speaking children use null subjects systematically, it has been proposed that they begin with the initial parameter setting allowing null arguments (NAs), and must change this setting on the basis of linguistic evidence that adult English prohibits NAs. A recent proposal suggests that the licensing and identification of NAs used by English-speaking children is like that used in adult Chinese. This predicts that young Chinese- and English-speaking children should exhibit parallel performance in their use of NAs. This study investigated this prediction using an elicited production task with both Chinese- and English-speaking children. Although the hypothesis that early English allows null *subjects* was upheld, the evidence is against the claim that early English is a discourse-oriented language like Chinese: while the Chinese children systematically used null objects, the American children did not. An alternative analysis of the use of null arguments is suggested.

1. INTRODUCTION

1.1 The Null Subject Phenomenon in Early Child Language

The null subject phenomenon, i.e., the frequent absence of lexical subjects, is one of the most noticeable characteristics of early child language. The following (non-imperative) English sentences (1a) and (2a), spoken by children aged from 1;8 to 2;5 (cited by Hyams, 1983), are examples of this phenomenon.

(1) a.	Read bear book Ride truck Want look a man	(1) b.	Kathryn read this Gia ride bike I want take this off
(2) a.	Outside cold No morning Yes, is toys in there	(2) b.	('It's cold outside') ('It's not morning') ('Yes, there are toys in there')

In the examples in (1a) the subject, though not phonologically specified, has a definite reference which can be readily inferred from context. Since sentences with null subjects like those in (1a) co-occur with sentences like those in (1b), which do have lexical subjects, it is not likely that the missing subjects in (1a) can be attributed to a performance constraint on sentence length. A further characteristic of children's speech at this age is illustrated by the examples in (2a). In these examples the unexpressed subject is an expletive, as shown by the 'translations' of these sentences in (2b). However, according to Hyams, children at this age do not produce sentences such as (2b).

Additional studies of children's early use of subjectless sentences are found with both languages which do allow null subjects and those which do not, such as Italian (Hyams, 1986),

This study was supported in part by NIH grants DC-00403 to Dr. Catherine Best and DC000183 to Dr. Diane Lillo-Martin, both at Haskins Laboratories.

We would like to thank the Chinese children and their families, and the teachers and the American children at the Child Development Lab of the University of Connecticut, for their help and participation in this experiment. Without them, it would have been impossible to have accomplished this study. We are grateful to the professors and students in Linguistics Department at the University of Connecticut who have given us valuable comments at various times. We would also like to thank Mr. Philip Chagnon at Haskins Laboratories, who helped us greatly in making the wood dolls for the experiment. We would also like to express our gratitude to the reviewers for their helpful comments and suggestions.

German (Clahsen, 1989; Weissenborn, in press), French (Weissenborn, in press), and American Sign Language (Lillo-Martin, 1986, 1991). In all of these studies, it has been found that at an early age children use subjectless sentences like the ones illustrated in English above.

The search for an explanation of children's early use of subjectless sentences can be related to studies of adult languages which permit such sentences as grammatically acceptable, by comparison to those which do not. In the next section, we review some characteristics of the null subject phenomenon in adult languages (since we include null objects as well as null subjects, the term has been generalized to 'null arguments'), and one proposal for the grammatical mechanisms underlying this phenomenon. We will then turn to a proposal accounting for children's use of null subject sentences which appeals to this analysis of adult language.

1.2 The Null Argument Phenomenon in Adult Languages

The null argument phenomenon is a well-known characteristic of adult languages such as Spanish, Italian and Chinese. Examples from these languages are given in (3). The English counterparts to these sentences require overt subjects.

In these so-called 'pro-drop' languages, the expletive elements equivalent to English it and there are also phonologically null, as illustrated in (4) (Italian, from Hyams, 1983), and (5) (Chinese).[2]

In adult Chinese, the expletive element equivalent to English it can be phonologically null as in Spanish or Italian, as illustrated above (5a, b, c).[3] Alternatively, a non-expletive subject can be found in any of these sentence types, illustrated in (6a, b, c).

(3) a. Mangia come una bestia. (Italian; Hyams, 1983)
'(He/she) eats like a beast.'

b. Come como una bestia. (Spanish; Hyams, 1986)
'(He/she) eats like a beast.'

c. [e] lái-le. (Chinese; Huang, 1982)
come-ASP[1]
'(He/she) came.'

(4) a. Sembra che Gianni sia matto.
'(It) seems that John is crazy.'

b. Piove oggi.
'(It) rains today.'

(5) a. [e] Xiàyǔ-le.
(It) rain-ASP
'(It) is raining.'

b. [e] Yào xiàyǔ-le.
(It) going to rain-ASP
'(It) is going to rain.'

c. [e_i] Kànshàngqù [e_i] yao xiàyǔ-le.
(It) seem (it) going to rain-ASP
'(It) seems that (it) is going to rain.'

(6) a. Tiān xiàyǔ-le.
sky rain-ASP
Lit., 'The sky is raining.'

b. Tiān yào xiàyǔ-le.
sky going to rain-ASP
Lit., 'The sky is going to rain.'

c. Tiān$_i$ kànshàngqù [e$_i$] yào xiàyǔ-le.
sky seem going to rain-ASP
Lit., 'The sky seems to be going to rain.'

How can one account for the occurrence of null arguments in these languages, compared to languages which prohibit null arguments, such as English? Jaeggli and Safir (1989) proposed the following Null Subject Parameter, stated in (7), as a principle of Universal Grammar (UG) to make this distinction.

(7) *The Null Subject Parameter*
Null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms.
(Jaeggli and Safir, 1989, p. 29).

According to Jaeggli and Safir, a morphological paradigm is uniform if all its forms are morphologically complex or none of them are. For example, the Italian inflectional paradigm consists entirely of morphologically complex forms, hence null subjects are allowed; in Chinese, no forms are morphologically complex, hence null subjects are allowed here too. In the case of English, however, morphologically complex forms such as *walks, walked, walking,* coexist with morphologically simple forms, such as *walk.* Thus English is a 'mixed' system and null subjects are prohibited.

The Null Subject Parameter stated in (7) tells us when a null subject is possible. However, Jaeggli and Safir (following others such as Rizzi, 1986) also propose that a null subject can occur only when its referential value can be recovered. They propose three mechanisms for the identification of null arguments: (i) local AG(reement) including a tense feature, (ii) a c-commanding nominal, or (iii) a Topic. Failure to satisfy either of the two necessary and sufficient conditions, namely, a morphologically uniform paradigm and a recoverable referential value for the thematic null subject, will result in the prohibition of null subjects in a language. Although the use of null arguments thus requires two conditions to be met, for ease of exposition we will refer to a Null Subject (or Argument) parameter with settings [+/-pro-drop]. (This also enables us to be neutral with respect to other analyses of the null argument phenomenon.)

The use of local AG to identify the reference of a null argument follows from numerous reports in the literature linking null arguments with 'rich' agreement. Early reports were confined to languages with only subject-verb agreement (such as Italian, discussed in Rizzi, 1982); these languages allow null arguments identified by agreement only in subject position. Later studies (such as McCloskey and Hale's 1984 work on Irish) have demonstrated that languages with other types of agreement often display null arguments in other positions. Jaeggli and Safir add the condition that a tense feature must be present in order to account for the lack of null arguments in German and other V2 (verb-second) languages. The null arguments which are identified by AG are considered to be members of the empty category *pro*, [+pronominal, -anaphoric].

The use of a Topic to identify null subjects follows from Huang's (1984; 1989) proposal. Huang distinguishes "discourse-oriented" languages from "sentence-oriented" languages. The "discourse-oriented" languages, like Chinese, have a rule of "topic-chaining" by which the discourse topic is grammatically linked to a null sentence topic which in turn identifies a null

argument. This null argument is a variable left from the movement of the empty topic to sentence-topic position. According to Huang, a topic may bind a variable in either subject or object position. These two kinds of null arguments are illustrated in (8).

In addition, there is a third method of identifying null arguments which results in a subject/object asymmetry. Because a c-commanding NP can also be an identifier, in languages like Chinese a null pronominal (*pro*) may be found in embedded subject position, as in (9a), but not in object position, as in (9b). This distinction is found because the empty embedded subject can be identified by the matrix subject; it functions grammatically like a pronominal rather than a variable. However, the empty object cannot be identified by the matrix subject, since identification has to be by the *closest* nominal element.[4] Thus, empty objects can only be identified by an empty topic, indicated by OP in (10).

To summarize, Jaeggli and Safir proposed that the difference between the grammar of pro drop languages such as Italian versus those such as Chinese is the method of identification of the null argument. This is illustrated in (11).

(8) a. Discourse Topic$_i$ [$_{S'}$ topic$_i$ [$_S$ [e$_i$] INFL lái-le]]
come-ASP
'(He) came.'
(Huang, 1984)

b. Discourse Topic$_i$ [$_{S'}$ topic$_i$ [$_S$ wǒ INFL [méi kànjiàn [e$_i$]]]]
I not see (him$_i$)
'I did not see (him).'

(9) a. Zhāngsān$_i$, tā$_i$ shū o[e$_i$] méi kànjiàn Lǐsì (Huang, 1989)
Zhangsan he say no see Lisi
'Zhangsan$_i$, he$_i$ said that (he$_i$) didn't see Lisi.'

b. *Zhāngsān$_i$, tā$_i$ shūo Lǐsì méi kànjiàn [e$_i$]
Zhangsan he say Lisi no see
'Zhangsan$_i$, he$_i$ said that Lisi didn't see (him$_i$).'

(10) [OP$_i$ [Zhāngsān$_j$ shūo [Lǐsì$_k$ kànjiàn [e$_i$] le]]]
Zhangsan say Lisi see ASP
'Zhangsan$_j$ said that Lisi$_k$ saw him$_{i/*j/*k}$.'

(11) a. [$_S$ pro$_i$ [$_{INFL}$ AG$_i$/Tense] ......]
(identification by AG, Italian)

b. Discourse Topic$_i$ [topic$_i$ [$_S$ t$_i$ [INFL] ......]
(identification by a discourse topic, Chinese)

c. Subject$_i$ verb [$_S$ pro$_i$ VP]
(identification by a c-commanding NP, Chinese)

1.3 Null Subjects in Children's Grammars

From the above, it may be seen that 'Early' English resembles a pro-drop language in three respects. First, lexical subjects are optional; second, the subject has definite reference even when phonologically null (except in the case of null expletives); and third, lexical expletives are absent (Hyams, 1983; 1989).

How can one account for the development an English-learning child has to undergo in order to arrive ultimately at a steady state grammar so as to speak the right type of English? A recent analysis by Hyams (in press; Jaeggli & Hyams, 1987), following the analysis of null subjects in adult languages by Jaeggli and Safir (1989) discussed above, proposed that the early grammar, like adult grammars, is constrained by the Null Subject Parameter cited above. That is, the early grammar satisfies the requirement of morphological uniformity and the requirement that null arguments be properly identified.

Hyams argues that English-speaking children begin speaking a Chinese-like language, i.e., a discourse-oriented language. Under the child's initial analysis, English is morphologically uniform with uniformly simple forms. Hyams takes children's verb productions, which at this time are generally *not* inflected, as evidence for this position. She further proposes that young English-speaking children use null topics to identify the reference of their null subjects. The child will then need to learn that English is not a 'Discourse Oriented' language in order to properly exclude null subjects.

In the case of Italian-speaking children, Hyams proposes that their early empty subjects are identified by AG(reement), as is the case in adult Italian. She proposes this early correct null subject use since Italian speaking children acquire the inflectional system fairly early. Thus, for these children resetting of the null subject parameter is not required.

One potential problem for Hyams' analysis is that one would expect that a discourse-oriented child language should have both null subjects and null objects, since under topic identification the null subject and null object phenomena are grammatically equivalent. However, according to the data she reviewed, Hyams claimed that English-speaking children do not use null objects. In order to account for this, Hyams thus proposed, following Roeper, Rooth, Mallis, and Akiyama (1984),[6] that in the early grammar, the inventory of null elements includes *pro*, but not variables. This hypothesis would predict a null subject/null object asymmetry. Since null objects can only be variables, under this hypothesis null objects would not be allowed in the early grammar until some later point when variables mature. In order for this account to hold, Hyams must depart from Huang's analyses of Chinese, and suggest that matrix empty subjects as well as embedded empty subjects can be *pro*, although only embedded empty subjects can be identified by a c-commanding NP. Hyams says that matrix empty subject *pro*s are identified by a discourse topic.

According to Hyams' hypothesis, Chinese-speaking children, who will ultimately acquire a real discourse-oriented language, should first exhibit the same null subject/null object asymmetry as English-speaking children, and they should not produce null object structures until the point when they develop variables. Hyams' hypothesis would also predict one of two null subject-object asymmetries for English-speaking children. On the one hand, if they have not yet reset the Null Subject Parameter by the time that they acquire variables, then they will produce only null subjects early on, but will later include null objects as well once they have developed variables. On the other hand, if the English-speaking children have reset the null subject parameter *before* they develop variables, they will never use null objects. Thus, knowing when English- and Chinese-learning children use null subjects and objects compared to when they develop variables is important for evaluating Hyams' proposal.

The evidence regarding the timing of use of variables versus resetting the null subject parameter is not wholly consistent with Hyams' approach. Roeper (1986) gives evidence that children have some uses of variables by age three to four years. All of his evidence for the use of *pro*s rather than variables with wh-questions occurs with older children (ages 8 to 10) and long-distance questions. However, his proposal that children use *pro*s instead of variables even at this later age can also be questioned, given new evidence regarding children's very early comprehension and production of wh-questions and strong crossover constructions (see Thornton, 1990). We therefore used the production and comprehension of wh-questions in the study reported here as evidence for the existence of variables in children's grammars.

The timing of the use of null subjects is easier to determine. The acquisition data Hyams used to support her hypothesis indicate that the restructuring of the Null Subject Parameter takes place around 26 to 28 months. If Hyams' proposal

that young children do not have variables is true, then we will not expect to see any null objects in the production of English-speaking children, since the restructuring takes place prior to the development of variables; and of course a clear decline in their use of null subjects should appear following the resetting of the NA parameter around 2-1/2 years. However, if there is evidence that children do have variables while they still use null subjects (indicating that the resetting of the NA parameter has not yet taken place), then they will be expected to use null objects too, according to Hyams' account.

In order to more fully evaluate Jaeggli and Hyams' proposals, we collected data on the acquisition of English and Chinese. The following experiment was designed to answer some relevant questions about Hyams' hypothesis through first-hand acquisition data. The questions we addressed include the following:

i. Is a null subject/null object asymmetry exhibited in child Chinese and child English? If so, is it equivalent for the two groups?

ii. If child Chinese or child English does exhibit null objects, do we have evidence that variables coexist with null objects? The emergence of *wh*-questions will be taken as evidence of acquisition of variables.

iii. Can the presence of lexical expletives be taken by American children as evidence that English is not [+pro-drop]? The use of overt versus null expletives will be examined to address this question.

iv. What does the developmental pattern look like, as far as the null subject and null object phenomena are concerned, in terms of the parameterized theory of UG?

v. What is the influence of linguistic environment during development of early grammar between ages 2 - 4-1/2?

2. Method

2.1 Subjects

2.1.1 Chinese and American children. Nine Chinese children, 4 female and 5 male, aged from 2;0 to 4;6, participated in the experiment. All of them were learning some variety of Mandarin Chinese as their first language. Their parents were graduate students from either mainland China or Taiwan, studying in the United States. Nine English-speaking children, 5 female and 4 male, aged from 2;5 to 4;5, were also tested using the same procedure. Their parents were members of the University community. All the subjects had normal hearing. There were no recorded developmental delays of any sort. Subject characteristics are given in Appendix 1.

2.1.2 Chinese adult controls. Nine Chinese-speaking female adults participated in the experiment. They were all born in mainland China or Taiwan, speaking some variety of Mandarin Chinese. They were the mothers of the Chinese child subjects.

2.2 Procedure

2.2.1 Controlled production data collection. This part of the experiment was carried out in the experimenter's home for the Chinese children, and in the observation room at a day care center for the English-speaking children. There were two story books used. One was a story book designed by the experimenter (QW) about the daily life of a little boy named Baldy (who had no hair). A doll house with dolls and furniture corresponding to the settings and characters in the book was used to familiarize the subject with the main character. Another story used was a pop-up book, "The Three Little Pigs." The testing was carried out after the experimenter played with the child subject a number of times and established rapport. The subject's task was to tell the experimenter the story. For the first story, the experimenter and the subject played with the doll house and dolls. Next, the subject was asked if he or she wanted to read a book about Baldy and then to tell a story about him. The answer was invariably positive. The entire procedure was audio recorded. All interaction with the Chinese-speaking children was conducted in Mandarin; that with the English-speaking children was in English.[6]

2.2.2 Eliciting expletive structures. In this part of the experiment, a number of pictures were displayed to the child subject and then he or she was asked to tell what happened in the pictures. This part of the experiment was designed to elicit expletive structures for the English-speaking children and to compare their productions to those produced by the Chinese-speaking children under the same situation.

2.2.3 Adult controls. The Chinese adult subjects were asked to tell the stories and talk about the pictures, while pretending that they were talking to their own child. The testing was conducted in the subjects' home without their child or the experimenter present. The testing materials were identical to those prepared for the child subjects. The whole procedure was audio taped.

2.3 Data reduction

i. The mean percentage of sentences with null subjects for each speaker was calculated based on the ratios of the sentences with null subjects to

the total number of sentences produced when telling the two stories. These ratios were averaged over the total number of subjects in each language group, over each age level (2-, 3-, and 4-year olds), and over each MLU level (3.5, 4.5, 5.25) separately. The standard error of the means (s.e.) was also calculated.[7]

ii. The mean percentage of sentences with null objects was calculated using a similar method. The ratio was the total number of sentences with an underlying structure of SVO to the total number of sentences produced with a null object. For the Chinese data, in addition to this criterion, any two-morpheme compounds which have been identified as a *word* by the authoritative dictionary—Xiàndài Hànyǔ Cídiǎn (Modern Chinese Dictionary) (Institute of Linguistics, Chinese Academy of Sciences, 1973)—were not included, even if they had the V+O formation. For example, (12a) was identified as a single word, so it was excluded; but (12b) was counted because it was not identified as a single word. The reason for this constraint is that it is generally agreed among Chinese linguists that a verb+complement compound is not equal to the structure of V+O; unlike the latter, the former is already in its minimal construction and is not divisible; therefore, these two types of words are analyzed differently.

(12) a. xǐ zǎo
wash bath
'take a bath'

(12) b. xǐ shǒu
wash hands
'wash hands'

iii. The MLU for child subjects in both languages was calculated, using the productions made for the stories, according to the method in Brown (1973).

iv. A second measure of the mean percentage of sentences with null subjects for English-speaking children was also calculated in the same way, excluding the sentences with null subjects using a gerund or to-infinitive. The reason for this exclusion is that given the discourse, these kinds of sentences are also allowed in the adult grammar of English. This second measure is labelled 'adjusted' in the figures.

v. The data gathered from testing the expletive structures was excluded from the calculation of the mean percentages. This part of the data was only evaluated for structural differences among the three testing populations. No quantitative analysis was involved.

vi. The children's comprehension and spontaneous productions of wh-questions during the course of the study were evaluated, for the purpose of determining their use of variables.

3. Results

3.1 An Overall View of the Results (for details see Appendices 2 and 3)

3.1.1 Null subjects. From Figure 1, it may be seen that there is a noticeable difference between the mean percentages of sentences with null subjects produced by Chinese child subjects and that by American child subjects at 2 - 4-1/2 years. Examples for such sentences are (13a,b) for the Chinese child subjects, and (14a,b,c) for American child subjects.

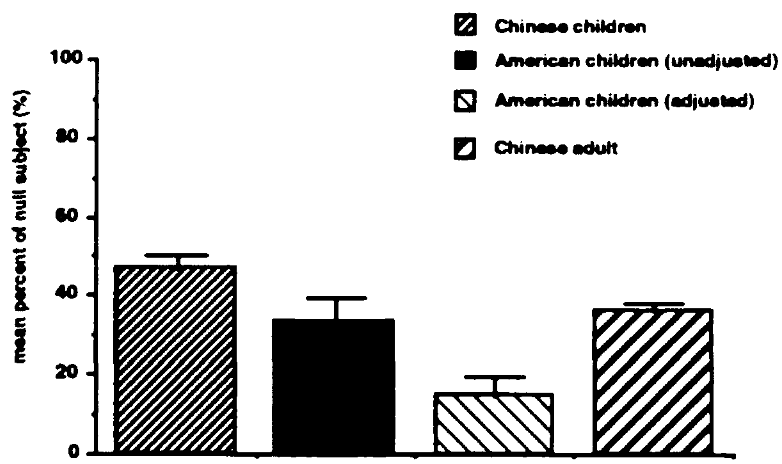

Figure 1. **Mean percentage of sentences with null subjects produced by Chinese and American children and Chinese adults.**

(13) a. Zhè húang wáwá tiàotiào. [e] shūai. [e] shūai dǎo le.
this yellow baby jump fall fall down ASP
'This yellow baby jumped. (He) fell. (He) fell down.'
(ZY, 2;0)

b. [e] wán shāshā ne. [e] zāng. [e] xǐ zǎozǎo ne.
play sand NE dirty take bath NE
'(He) is playing with sand. (He) is dirty. (He) is taking a bath.'
(AN, 2;3)

(14) a. [e] brush her hair. [e] brush hair.
'(She's) brushing her hair. (She's) brushing (her) hair.'

[e] fighting like that, bang!
'(They're) fighting like that, bang!'

[e] playing. They all bent. [e] are playing.
'(They are) playing. They (are) all bent. (They) are playing.'
(AR,2;5)

b. He got in there. [e] fell down.
'He got in there. (He) fell down.'
(DS,2;10)

c. [e] jumping. [e] fell. They fell down. [e] sleeping.
'(They're) jumping. (They) fell. They fell down.
(They're) sleeping.'
(SP,4;2)

The mean percentage of sentences with null subjects produced by Chinese children is 46.54% (s.e. = 3.78); while for the American children, it is 33.11% (s.e. = 6.12). The Chinese adults produced sentences with null subjects 36.13% of the time. Given that Chinese is a pro-drop language, all the sentences with null subjects produced by the Chinese children are considered grammatical, with the reference of the null subject determined by the discourse topic. Although English is not a pro-drop language, some of the sentences with null subjects produced by American children, i.e., sentences with null subjects but using infinitives or gerunds rather than a full verb, can be judged as pragmatically acceptable in the given context in which they were produced. If we exclude these sentences from our count of sentences with null subjects produced by American children, the mean percentage drops to 14.58% (s.e. = 5.03). Comparing this adjusted mean percentage, 14.58%, with the mean percentage of Chinese children, 46.54%, and that of Chinese adults, 36.13% [one way ANOVA omnibus $F(2, 24)=17.80$, $p=.0001$], it is clear that Chinese children are dropping their subjects at a much higher rate than American children, and even a bit higher than the rate of the Chinese adults. The differences between the American children and the Chinese children, and between the American children and the Chinese adults, are both significant by Scheffé's tests [$F(1,24)=31.96$, $p=.0001$, and $F(1, 24)=21.55$, $p=.0025$ respectively]; the difference between the Chinese children and the Chinese adults is not significant. Even still, it is clear that American children do drop subjects a relevant amount of the time.

For both groups of children, the null subject was sometimes clearly related to an antecedent from the discourse as shown in examples (15, Chinese) and (16, English). In other cases, the referent of the null subject was not previously mentioned in the discourse, although it was usually understandable from the context; often, it was part of the pictures the children were describing. Some examples of this type are given in (17, Chinese) and (18, English).

(15) a. Xiǎo zhūzhū zhǔ tāngtāng.
little piggy boil soup
'Little pig makes soup.'
[e] zhǔ tāngtāng.
(He) boil soup
'He makes soup.'
(WW, 2;5)

b. Dà yě láng$_i$ zài zhèlǐ tōu kàn.
Big wild wolf$_i$ ASP here secretly look
'The big wild wolf is here peeping secretly.'
[e$_i$] zài kàn xiǎo zhū.
(It$_i$) ASP look little pig
'It is looking at the little pig.'
(HE, 3;1)

(16) a. Look at this bad wolf. He got in there. [e] fell down.
'Look at this bad wolf. He got in there. (He) fell down.'
(DS, 2;10)

b. The big bad wolf coming again and bang the door. [e] want to blow the house and the house is down.
'The big bad wolf (is) coming again and bang the door. (He) wants to blow the house and the house is down.'
(SR, 2;8)

(17) [e] kàn jìngjìng. [e] méi chūan xiéxié
(He) look mirror (He) not wear shoe
'He is looking in a mirror. He didn't wear shoes.'
[e] méi chūan wàwà.
(He) not wear sock
'He didn't wear socks.'
(ZY, 2;0)

(18) [e] jump up. [e] jump in bed. [e] fall down.
'(He) jumped up. (He) jumped in bed. (He) fell down.'
(AR, 2;5)

Although both Chinese- and English-speaking children thus produced null subjects in a somewhat similar fashion, we believe this does not necessary show that they use the same mechanism in identifying and licensing the null subjects. We will return for further discussion of this point.

3.1.2 Null objects. From Figure 2, we may see that there is a considerable difference between the mean percentages of sentences with null objects produced by Chinese child subjects, which is 22.53% (s.e.=1.76), or by Chinese adults, 10.3% (s.e.=1.58), and that by American child subjects, which is 3.75% (s.e.=1.31), [one way ANOVA omnibus $F(2, 24)=37.21$, $p=.0001$]. Here, the differences between the American children and the Chinese children, the American children and the Chinese adults, and the Chinese children and the Chinese adults are all significant by Scheffe's tests [$F(1,24)=18.781$, $p=.0001$, $F(1, 24)=6.549$, $p=.0237$, and $F(1,24)=12.232$, $p=.0001$, respectively]. With the Chinese children, only 27.59% of the total sentences with null objects are ungrammatical. The grammaticality of the Chinese object-drop sentences (i.e., whether the null object was used properly) was judged with respect to the context in which the sentence in question was actually produced. For the American children, 100% of the sentences with null objects were ungrammatical. Examples are given in (19) for Chinese child subjects, and (20) for American child subjects.

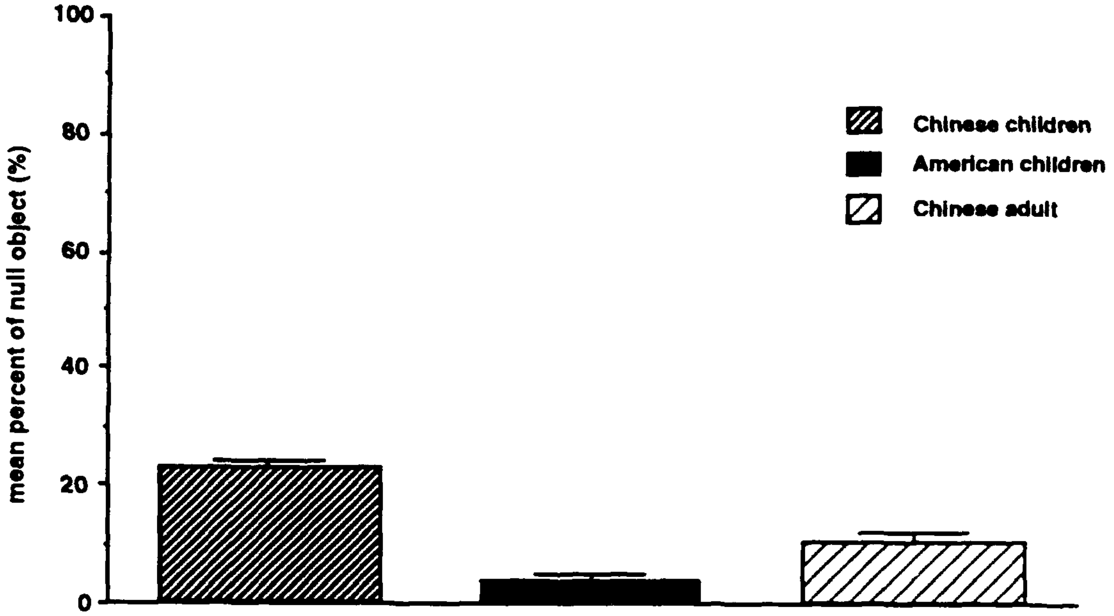

Figure 2. **Mean percentage of sentences with null objects produced by Chinese and American children and Chinese adults.**

(19) a. *Ou, láng lái chī [e]. (ungrammatical)
oh, wolf come eat (it=pig)
'Oh, the wolf came to eat (the pig).'
(ZY,2;0)

b. *Tāmén yào qiù gài [e]. (ungrammatical)
they going to build (it=house)
'They are going to build (a house).'
(WW,2;5)

c. [e] Zài kànkàn [e]. (grammatical)

(He=wolf) again look look (it=pig)
'(He) had another look at (the pig).'
(ZY,2;0)

d. [ei] chīwán [ej], (grammatical)
(He=wolf) eat finish (it=pig)
'After (he) finished eating (the pig),'
lǎoláng dùzi jiòu biàn dà le.
old wolf belly then become big ASP
'the old wolf's belly became big.'
(LX,3;4)

(20) a. *Look at [ei]. [ei] go a little higher (ungrammatical)
'Look at (him). (He) goes up a little higher.'
(DS,2;10)

b. *The other little pigs worry about [e]. (ungrammatical)
'The other little pigs worry about (him).'
(ER,3;8)

3.1.3 Null subject/null object asymmetry. Comparing Figure 1 with Figure 2, it may be seen that the null subject/null object asymmetry is not unique to the Chinese children. The ratio of the mean percentage of sentences with null objects to those with null subjects is 0.48, 0.23, and 0.24 for Chinese children, Chinese adults, and American children, respectively. If we recalculate the ratio for the Chinese children, excluding the ungrammatical sentences as in example (19a and b), (which may be considered as errors), the ratio decreases from 0.48 to 0.29. If we do the same thing for the English children, considering their small percentage of object-dropping (3.57), which was ungrammatical, as errors, the ratio of course becomes zero.

The amount of null object use by the Chinese adults is surprisingly low. However, it is important to note that we believe the ratio for Chinese adults would be higher than the rate we obtained if the data had been collected in an Adult-to-adult conversational situation, where most object dropping takes place, rather than in children's storytelling. Because of this discrepancy, we conducted a follow-up study with Chinese adults.

In the follow-up study, five Chinese-speaking adults were interviewed by the experimenter in an adult-to-adult conversational setting. These adults were all women who had recently given birth to their first child. The interviews took place in the subjects' homes, and consisted of several parts. First, the subjects were asked to tell their child two stories as a warming up. Then, they engaged in conversation with the experimenter. The conversations all included the same three topics of discussion: the woman's pregnancy and childbirth, her own lifestyle, and the growth and behavior of her child. The interviews were tape-recorded. Only the conversations were transcribed and scored according to the same procedures discussed previously for the initial study. The percentages of null subject and null object used by each speaker in this study are illustrated in Figure 3, and more detailed information is given in Appendix 4.

As this Figure clearly shows, a subject-object asymmetry remains for the adult subjects, but the overall percentage of null object use increases greatly. Both of these facts are important for comparison with the children's utterances. In the follow-up study, the average object drop is 40.1% (s.e.=1.77), while the average subject drop is 45.6% (s.e.=2.42). Although the amount of object drop is much higher than in the initial study (10.30%), the difference between the subject-drop and the object-drop is significant by a 2-tail paired t-test (t=4.073, p=0.015). Some examples of the adults' utterances with subject and/or object drop are given in (21) and (22).

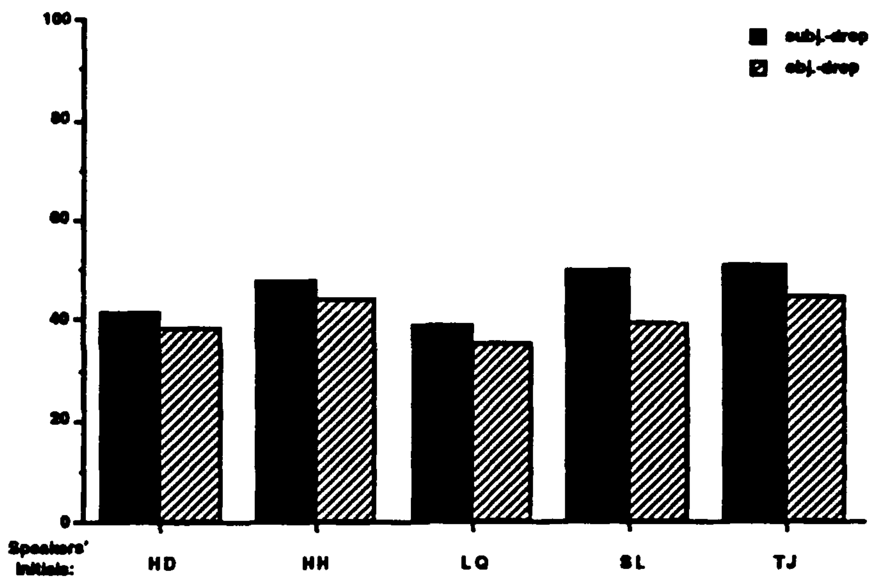

Figure 3. **Percentage of sentences with null arguments produced by Chinese adults in the follow-up study.**

(21) Tā$_i$ jiòu hē diǎn niúnǎi$_j$ ma. [e$_i$] yě hē [e$_j$] bù duō.
He$_i$ only drink little milk MA. (He$_i$) yet drink (it$_j$) not much.
'He only drinks a little milk. (He) does not drink (it) much.'

[e$_i$] zài hē diǎn gǔozhī. [e$_i$] chī diǎn shǔiguǒ.
(He$_i$) also drink a little bit juice. (He$_i$) eat a little bit fruit
'(He) also drinks a little bit of juice. (He) eats a little bit of fruit.'
(LQ)

(22) Tā$_i$ yiòu ài kàn diànshǐ$_j$.
She$_i$ especially like watch TV$_j$
'She especially likes to watch TV.'

Wǒ$_k$ jiòu pà tā$_i$ bǎ yǎnjīng kàn huài-LE.
I$_k$ so afraid she$_i$ BA eye watch bad-ASP.
'I was so afraid that she might damage her eyesight.'

[e$_k$] yī tiān bú ràng tā$_i$ kàn nàmo jiǒu [e$_j$].
(I$_k$) a day not let her watch that long (it$_j$).
'I do not let her watch (it) for long in a day.'
(TJ)

3.2 Results Broken Down by Age and by MLU

In order to determine whether there is any relationship between the null subject /null object phenomena and the child's linguistic maturation, the results were recalculated according to the child's chronological age and the child's MLU level.

For the American children, the adjusted mean percentages of sentences with null subjects are 25.89%, 4.48%, and 13.39% for age group 2, 3, and 4 respectively. For the Chinese children, the mean percentages of sentences with null subjects are 55.73%, 45.65%, and 38.25% for these three age groups. Thus, in both languages, the proportion of subjectless sentences decreases over time. However, the American children seemed to make a surprising jump up in the use of null subjects by four-year-olds.

To investigate this further, the percentage of null subject sentences was recalculated on the basis of MLU. It was found that for the Chinese child subjects, MLU levels were in accordance with their chronological age groups; however for the American child subjects, the 2- and 3-year-old groups had MLU levels corresponding to their 2- and 3-year-old Chinese counterparts, but the 4-year olds had an MLU level corresponding to the Chinese 3-year-olds. Thus, the American 3- and 4-year-olds were grouped together in one MLU group for the comparison of null subjects across MLU.

Grouped by MLU, the American children produced subjectless sentences 25.89% of the time and 8.93% of the time for MLU level 3.51 (2-year olds) and 4.48 (3 and 4-year olds), respectively (see Figures 4 and 5). The difference between the Chinese and American first MLU groups (2-year-olds) is not statistically significant (t=2.209, p=.09), however, as can be seen in Appendix 1, this is essentially due to the youngest American subject (AR), who had a rate of subject drop comparable to that of his Chinese peers. The difference between the second MLU groups (Chinese 3-year-olds and American 3- and 4-year-olds) is significant by unpaired two-tail t-test (t=2.21, p =.0007). Clearly, the American children experience a sharp drop in their use of null subject sentences. The Chinese children, on the other hand, continue to use null subjects across the MLU groups (for the Chinese children, MLU groups are equivalent to age groups).

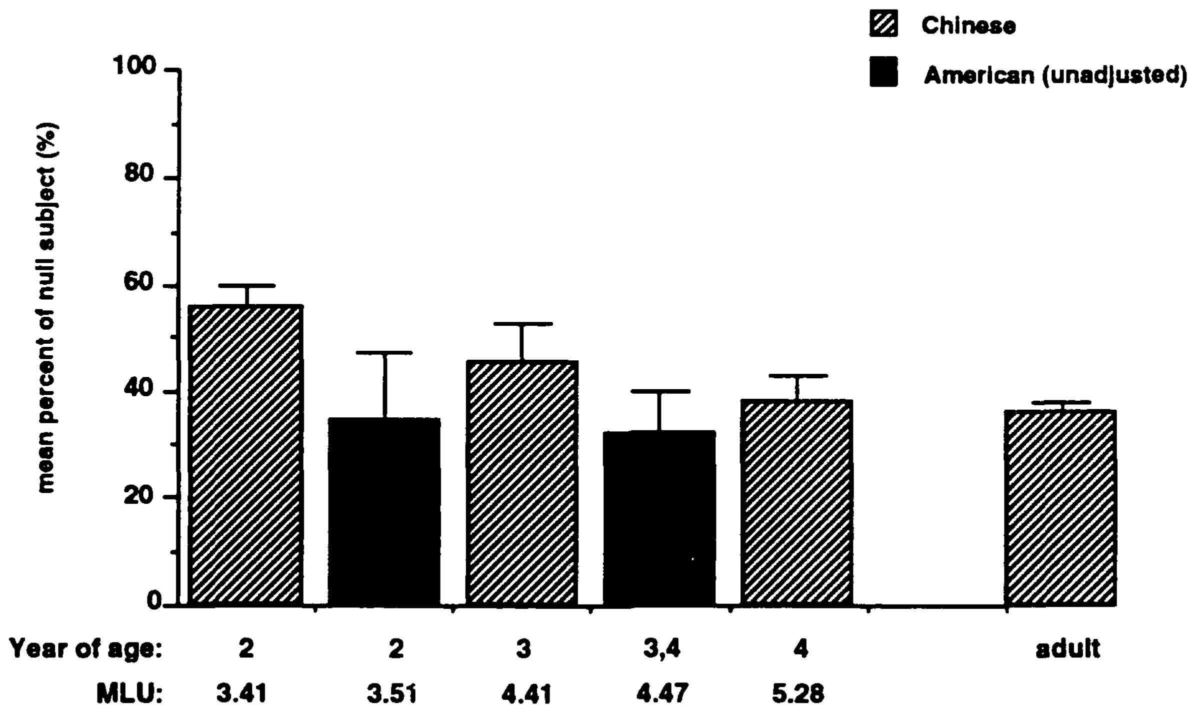

Figure 4. **Mean percentage of sentences with null subjects produced by Chinese and American children (by MLU, unadjusted) and Chinese adults.**

The pattern of use of missing objects is quite different (see Figures 5 and 6). Whether divided by age or by MLU group, the American children used missing objects much less frequently than null subjects. The two-year-olds MLU 3.51) used missing objects only 8.3% of the time, while the older children used essentially none. In contrast again, the Chinese children used null objects much more frequently than the American children. They averaged 20.2% to 26.0% null objects, with the figures increasing slightly over the age/MLU ranges.[8] Although the adults in the initial study produced far fewer null objects than the Chinese children, from the follow-up study we can see that the overall production of null objects by the children is approaching the level of use by adults in conversational settings.

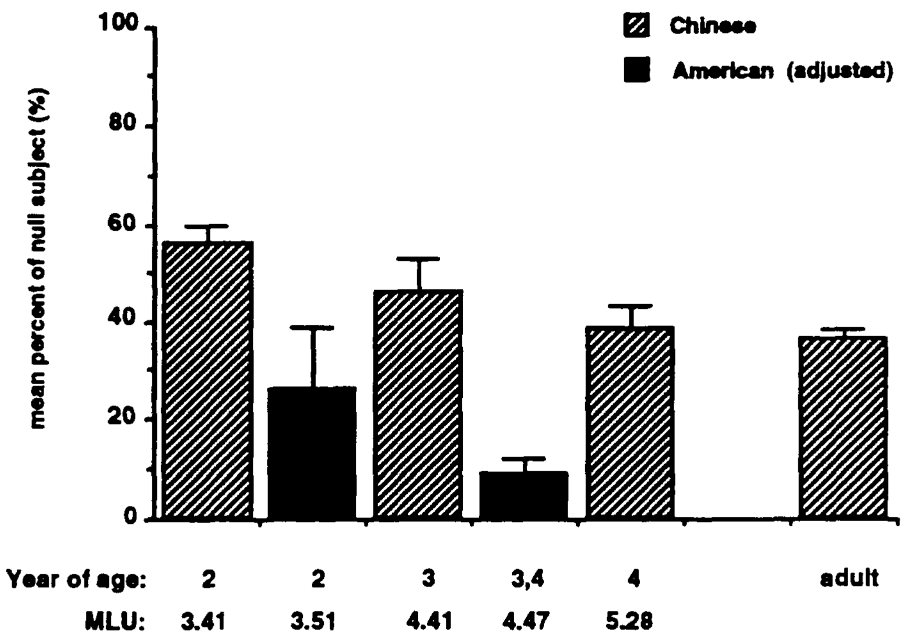

Figure 5. Mean percentage of sentences with null subjects produced by Chinese and American children (by MLU, adjusted) and Chinese adults.

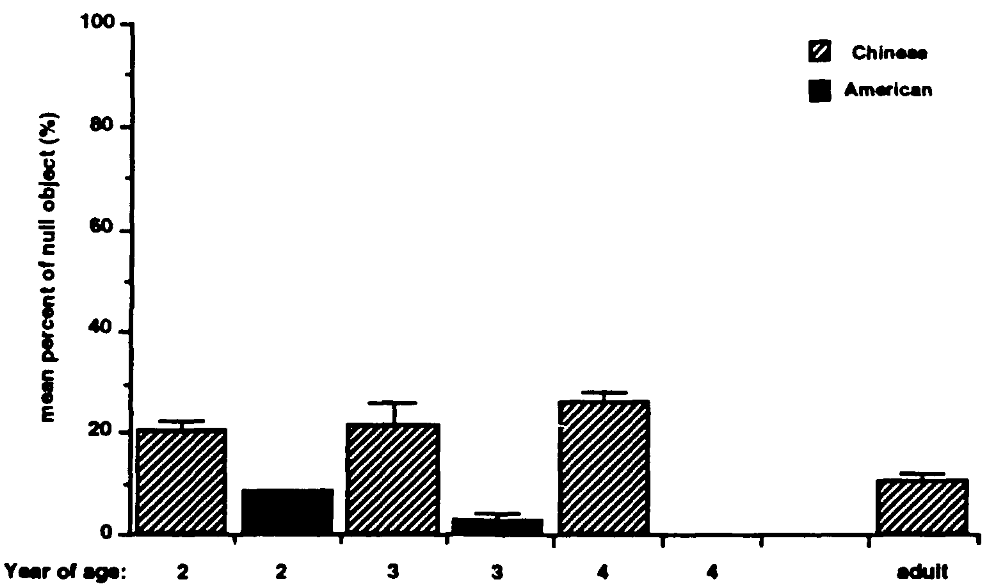

Figure 6. Mean percentage of sentences with null objects produced by Chinese and American children (by age) and Chinese adults.

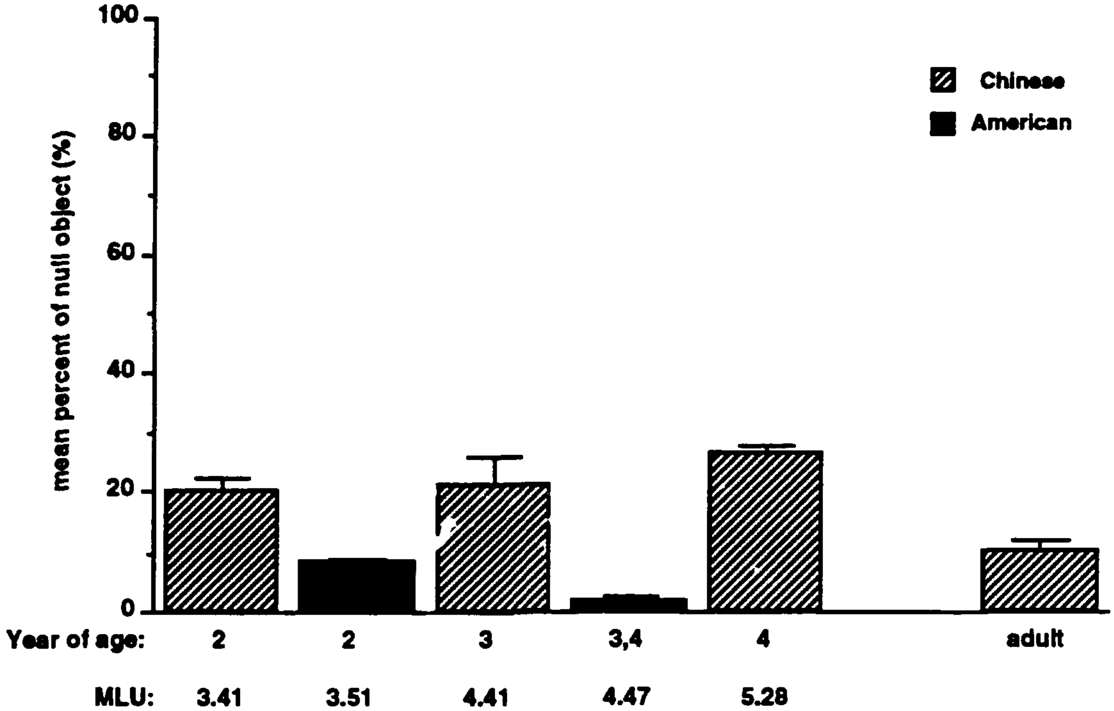

Figure 7. **Mean percentage of sentences with null objects produced by Chinese and American children (by MLU) and Chinese adults.**

The Chinese- and English-speaking children do not differ significantly in their use of null subjects at the earlier MLU stage tested: MLU level 3.5, but they do at the latter MLU stage: MLU level 4.5. These results provide strong evidence for pro-drop in younger English-speaking kids (MLU level 3.5). For the use of null objects, however, the two language groups differ significantly across all MLU levels. The differences in the use of null subjects and null objects by Chinese and American children indicate that the factors controlling the use of the two types of null arguments in the two groups are distinct. This is counter to the proposal by Jaeggli and Hyams (1987) which suggests that the two groups use null subjects for essentially the same reason.

3.3 Results of Eliciting Expletive Structures

In order to determine how the course of the development of expletive subjects interacts with the development of null versus overt subjects, children's productions of sentences calling for expletive subjects were examined. For the Chinese-speaking children, we examined whether they used a null subject as in (5) above, or a non-expletive lexical subject as in (6). For the English-speaking children, we examined whether they produced any lexical expletives, and further, whether there was any evidence that lexical and null expletives coexisted.

In general, there was no evidence of the Chinese children producing structures with overt non-expletive subjects, such as those in (6a, b, and c) above, even among the 4-year olds. The only structures they used in the weather conditions were those with null subjects, as in (5a and b). They did not use the structure as in (5c) either. The only exception occurred when they talked about a windy condition. In this case they either used a structure with a null subject as in (23), or they used 'fēng,' ('wind'), as an overt subject as in (24). The Chinese adults used all the structures as in (5) and (6). They also used 'fēng,' the word for 'wind,' in the same way as the Chinese children. The observed difference here between the Chinese children and the Chinese adults in their use of null subjects (as in 5a and b), and non-expletive

lexical subjects, (as in 6a and b), we believe, is due to a stylistic reason rather than a grammatical one. In fact, sentences in (5a and b) are more colloquial than those in (6a and b). However, it seems that the absence of the structure like that in (6c) from the data of the Chinese children is due to a grammatical reason. While the null subjects in (5a and b) can be interpreted as referential, the one in (6c) can not. The structure (as in 6c) requires the ability to raise the subject from the embedded clause to the matrix clause.

The American children had a different pattern. Except for the youngest one, (AR, 2;5), all the children showed some kind of evidence for the existence of expletive 'it' as in example (25). At the same time, however, they also used some null expletives as well, as shown in examples (25) and (26).

(23) [e] yào bǎ zhège guā diào,
(it=wind) want (BA) this blow down,

[e] hái yào bǎ zhège jě guā diào.
(it=wind) also want (BA) this too blow down.

'(Wind) wants to blow this down,
(it) also wants to blow this down too.'
(ML, 4;3)

(24) Xiànzài guā fēng-le. Fēng dōu tài dà-le,
now blow wind-ASP. Wind also too big-ASP
fángzi dōu chuī dǎo-le.
house also blow down-ASP
'The wind began blowing now. The wind was so big
that the house was blown down.'
(SK, 4;1)

(25) It is raining. (SR, 2;8)
It's very windy so the clothes are going up. (SR, 2;8)
It's rain. rain. They can't come out. (DS, 2;10)

(26) Snow. Raining (DS, 2;10)
No snow. (SR, 2;8)
Windy now. (EL, 3;6)
Raining. (AR, 2;5)

Hyams (1986) suggests that one piece of evidence that English-speaking children use to reset the null subject parameter to [-pro-drop] is the presence of overt expletives. Hyams argues that since *it* and *there* are not being used for pragmatic purposes (because they do not contribute to the meaning of the sentence), they must therefore be present for strictly grammatical reasons. Hence, lexical expletives could be used to trigger parameter resetting. Furthermore, as noted above, Hyams found that children use null expletives at the time they use null subjects. So the emergence of lexical expletives coincident with restructuring to [-pro-drop] is predicted.

However, as our data show, some children do use both overt and null expletives at the time when they are using null subjects. Hence, it seems that the presence of overt expletives in the input is not a type of triggering data for resetting the null subject parameter. But why do the children use overt expletives when they sanction null subjects? Lillo-Martin (1987) has given a reasonable solution for this puzzle. She suggests that children have misanalyzed the expletives, and instead interpret 'it' as referential, even in sentences like, 'It's raining.' Because they have the wrong analysis of 'it,' they don't have the overt expletive evidence that English is not [+pro-drop]. So at this point, one cannot assume that the time at which a child starts using overt expletives will be coincident with the correct setting for the null subject parameter.

3.4 Results on the Use of Structures Exhibiting Variables

In our data, both child language populations have shown some evidence for the existence of variables though the production of wh-movement (English), or the comprehension and production of wh-questions (Chinese). This can be seen in (27) and (28). These questions were produced and comprehended during the course of the experiment described above, at the same time as these children showed evidence of using null arguments.

One might claim, following Roeper et al. (1984), that the empty categories used in these constructions are *pros*, not variables. However, work by Thornton (1990) and Sarma (1991) suggests that children at least at 3 years do use variables rather than *pros* in these constructions, since they correctly produce long distance questions and obey the strong crossover constraint. Therefore, we will assume that the empty categories used in the wh-questions shown above are variables rather than *pros*. In any case, it is the difference between Chinese- and English-speaking children with respect to null objects, without a corresponding difference with respect to evidence for variables in the form of wh-questions, that is relevant to our discussion.

(27) a. What's that?
(AR, 2;5)

b. Who's that? Baldy? Baldy is playing with mud.
(SR, 2;8)

c. That's what I think he did.
(DR, 3;9)

(28) a. Experimenter: Shúi lái-le ?
Who came-ASP
'Who came?'

Child subject: Láng, Láng lái-le.
wolf, wolf came-ASP
'The wolf came.'
(ZY,2;0)

b. Experimenter: Dà hūi láng gàn shénmo lái-le ?
big grey wolf do what come-ASP
'Why did the big grey wolf come?'
Child subject: [e] Ná xiǎo zhū Ah .
(He) take little pig Ah!
'(He) came to take the little pig away, of course.'
(AN, 2;3)

c. Nà shì shémo? Nà shì shúi nòng de?
that is what? that is who did
'What is that?' 'Who did that?'
(WW, 2;5)

4. DISCUSSION: THE PARAMETERIZED THEORY OF UG AND LINGUISTIC EVIDENCE

A review of Figures 4 through 7 indicates the following:

i. At the earliest age tested, 2 years old or average MLU of 3.5, both Chinese and American children are using null subjects. The Chinese children are also using null objects. Although the American children do have a few sentences with null objects, the mean percentage of their sentences with null objects is only 3.57, so we will count these as errors; i.e., outside of the children's grammars.

ii. For the Chinese children, as their MLU increases, the mean percentage of sentences with null subjects decreases, and the mean percentage of sentences with null objects increases. By the MLU level of 5.28, their subject-dropping rate is very close to that of Chinese adults, and their object-dropping rate is approaching that of the adults in the follow-up study.

iii. For the American children, as their MLU increases, the mean percentage of sentences with null subjects (as well as sentences with null objects, which we are not counting as part of the children's grammar) decreases drastically, thus also coming in line with the corresponding adult grammar.

iv. At each MLU level, both mean percentages are much higher for the Chinese children than their American counterparts, although for the first MLU group (MLU level 3.5) the difference between the Chinese- and English-speaking children in their use of null subjects is not statistically significant.

How can the observation that as early as 2 years old both Chinese and American children are using null arguments be explained? It might be understandable that Chinese children do so because adult Chinese is a pro-drop language. But then why would the American children also do so, given that null arguments are not allowed in adult English? On the other hand, how can the observed differences between Chinese and American children in the null argument phenomena be explained along developmental lines?

If we adopt the idea that part of the formulation of UG is a system of parameters, and the initial setting for a particular parameter is the same for all children constrained by certain principles, then the observed phenomena can be explained. As discussed above in detail, the principles of UG may tell us when a null subject can occur and how it can be identified. The data we obtained support the hypothesis that English- and Chinese-speaking children at a very early age have a grammar which allows null subjects.

We are left, however, with three important questions for discussion. First, how strong is the asymmetry we found comparing subject and object dropping in English compared to Chinese, and how can it be accounted for by parameter theory? Second, how does the child who begins with an incorrect parameter setting make the change to the adult grammar? Third, how does the linguistic environment make an impact on this parameter resetting?

4.1 On the Subject/Object Asymmetry

Our data did not confirm Jaeggli and Hyams' hypothesis with respect to null objects. Instead, our data indicate that while the Chinese-speaking children used null objects from as early as 2 years old (the youngest age tested), the English-speaking children by and large did not use null objects. This returns us to the potential problem with Jaeggli and Hyams' account discussed above. If English-speaking children have a Chinese-type language as their initial parameter setting, then we would expect children learning both languages to progress similarly in terms of the use of null objects. However, this was not the case.

We do not think that the null subject/null object asymmetry we found in Chinese- and English-speaking children's use of null objects can be accounted for by the non-existence of variables in early grammar. Both the Chinese- and the English-speaking children provided evidence for the emergence of variables. According to Hyams' hypothesis, the English-speaking children in this situation should use null objects at least as productively as the Chinese-speaking children do, but our data show that they do not. The small percentage (3.57) is really within the error range. If the English-speaking children have reset their null argument parameter at this point, they should have stopped using both null subjects and objects. Our data show that this is not the case: they continued to use null subjects but essentially no null objects even though they had acquired variables. At the same time, the Chinese-speaking children (who showed the same kind of evidence of variables) did use null objects productively.

As an alternative to Jaeggli and Hyams' hypothesis, we propose that there is more than a single parameter controlling the use of null arguments (following Lillo-Martin, 1986; 1991). One parameter, which can be called the Discourse Oriented Parameter (DOP) (following Huang, 1984), permits languages with discourse oriented properties to have both null subjects and null objects. These null arguments can be one of two types. Most are variables identified by a Discourse Topic. In embedded subject position there is also the option of *pro*, identified by a c-commanding NP. These null arguments correspond straightforwardly to two of the identification options proposed by Jaeggli and Safir, given in (11b and c) above. For learnability reasons, assuming parameter setting takes place on the basis of positive evidence, we might expect that the initial setting of the DOP is [-DO]. If so, the performance of the Chinese-speaking children in our study indicates that resetting of the DOP to [+Discourse Oriented] can take place early. Since other characteristics of discourse oriented languages, such as topic-comment structures and discourse-bound anaphors, can serve as evidence for determining this parameter setting, it is reasonable to assume that the Chinese-speaking children have made this setting and produce null subjects and null objects in accord with this grammatical option.

The second part of our proposal is that null arguments in adult languages like Italian are due to a separate parameter, which we will call the Null Argument Parameter. This parameter permits null arguments when licensed by certain Case-assigning maximal categories, following Rizzi (1986). These null arguments are empty categories of the type *pro*, identified by the person, number-, and / or gender-features of the licensing category. Although subject-verb agreement is insufficient to license or identify null subjects in adult English, we take it that English-speaking children who use null subjects are doing so because of this parameter, rather than the DOP. The subject-object asymmetry is related to the cross-linguistic observation that object agreement is much less common than subject agreement; hence *pro* null objects are found in many fewer languages than *pro* null subjects. Children will universally posit an INFL category with the potential of being a licenser for empty subjects, but not for empty objects. Hence, universally children will begin with a null subject hypothesis. Changing the parameter setting to disallow null subjects will thus only take place after morphological agreement has been analyzed.

Other proposals have been made arguing that the null subject phenomenon in early English is due to performance factors rather than a grammatical parameter setting (e.g., Bloom, 1990; Gerken, 1990; Mazuka, Lust, Wakayama, and Snyder, 1986). Although these suggestions are worth considering, there is considerable cross-linguistic evidence to take the early null subject phenomenon as representing a grammatical stage. Performance accounts of the early null subject phenomenon do not make the same cross-linguistic predictions as grammatical accounts do. More cross-linguistic work can contribute to the resolution of this debate; but the data currently available support the grammatical account. For reviews of performance versus grammatical accounts, see Hyams and Wexler (1991) and Lillo-Martin (1991).

4.2 Parameter Resetting

The evidence is quite strong that both Chinese- and English-speaking children have a grammar which allows null subjects at an early age, since they were both using null subjects even at the age of 2 (examples 12a, and b and 13a, and b). For the Chinese children, since the adult language allows null arguments, no change will have to be made in their parameter setting. However, for the English-speaking children, a parameter will have to be reset on the basis of evidence for [-pro drop] from the linguistic environment. Our data shows that roughly between the age of 2 and 3 or MLU 3.5 to MLU 4.5, a drastic change has taken place in the English-speaking child's grammatical development. That is, during this transition the English-speaking children show a dramatic decline in the production of null subjects. It seems to be at this point that the parameter resetting has taken place.

How does this resetting occur? It is possible that the presence of overt expletives can be used as evidence that English is [-pro-drop], as discussed above. However, there is now some cross-linguistic data which indicates that the perfect correlation between overt expletives and [-pro-drop] which is needed for this kind of evidence does not exist (cf. Jaeggli & Hyams, 1987, Hyams, in press). Even if this positive evidence is unavailable, however, it is possible that indirect negative evidence can be used (Lasnik, 1989). For the English children, since the child's initial setting is also [+pro-drop], he would, like the Chinese children, expect to hear sentences with null subjects. When the child fails to hear sentences with null subjects in English, this will then be taken as indirect negative evidence that such sentences are not allowed in his language, hence, ungrammatical. The incorrect positive parameter will then be replaced by the correct negative setting [-pro-drop].

Note that our data do agree with some empirical data existing in the literature, which together may be taken as evidence for certain a priori, language-independent properties of early grammar hard-wired by parameters of UG. For instance, with our Chinese child subjects at MLU level 3.5, 20% of the transitive verb constructions were produced with null objects, which is very close to the 17% of the similar constructions obtained in Japanese children (Mazuka et al., 1986). Also, for the American child subjects, the mean percentage of sentences with null subjects (15%) is very close to the percentage found in Gerken's imitation study (19%, subjects' mean age was 2;3; Gerken, 1990). Further, the dramatic decrease in the mean percentage of sentences with null subjects observed in our American children between age 2 and 3 is consistent with Hyams' proposal of an inverse relationship between null subjects and the use of inflectional morphology. These studies all point to an initial [+pro-drop] setting, with resetting to [-pro-drop] for English-speaking children during the third year.

4.3 Effects of Linguistic Environment

What role does the linguistic environment play in this parameter-setting account of language development? Clearly, only data from the linguistic environment can trigger the resetting of a parameter, such as is needed for English-speaking children. However, the interaction between the child's initial setting of this null-subject parameter and the input of the child's linguistic environment seems to make itself felt even earlier and in more subtle ways than parameter resetting. Even the 2-year-olds we tested displayed a noticeable difference in the null subject/null object phenomena between the two testing populations. First of all, only the Chinese-speaking children used null objects to any extent. This, as we suggested, can be due to a different parameter from the one used for null subjects in English-speaking children; one that could possibly be set on the basis of entirely independent data.

A more extensive consideration of the role of the linguistic environment is called for if we take into account the proportions of null arguments used across the different age ranges in Chinese and English. Although the English-speaking children used null subjects frequently, they still used them less frequently than the Chinese children. In the case of null objects, we have suggested that the difference between English- and Chinese-speaking children is a difference related to their grammars: the Chinese-speaking children's grammars allow null objects, while the English-speaking children's grammars do not. However, we do not make the claim that the difference in the use of null subjects is a grammatical difference. This seems to be a prime example of an area where the force of the linguistic environment is felt. Furthermore, as they develop, the use of null arguments by the Chinese-speaking children approaches that of the adult subjects. For example, the Chinese adults produced sentences in which the null argument is interpreted by virtue of a discourse topic established several sentences earlier, as in example (22) above. The youngest children did not exhibit this kind of long distance topic chaining. The factors that control the pragmatically acceptable use

of null arguments (as opposed to their general grammaticality) will need to be learned by Chinese-speaking children, independent from the setting of grammatical parameters. This will be directly related to the linguistic environment.[9]

5. CONCLUSION

In general, this study has shown some support for the hypothesis that English-speaking children begin speaking a [+pro-drop] language. The specific hypothesis of Jaeggli and Hyams (1987), that early English is a Chinese-type language, received mixed support. Support in favor of Jaeggli and Hyams' proposal may be seen through the following points:

i. As early as 2 years old, which was the earliest age tested, the English-speaking children produced sentences with null subjects at 34.57%.

ii. The English-speaking children did display an asymmetry in the use of null subjects, compared to their very low incidence of null objects.

However, this data also throws Jaeggli and Hyams' (1987) theory into a dilemma. They use Roeper's (1986) proposal for the later development of variables in order to account for the proposed null subject/null object asymmetry. Our result shows that apart from the low level of null object errors, the English-speaking children never used any true null objects, consistent with Jaeggli and Hyams' analysis. However, we found this even after the children had developed variables (as indicated by production of Wh-questions). According to Hyams, the English-speaking children should have displayed null objects when they developed variables, or else they should have gone through the business of null argument parameter restructuring by this time, and displayed no null subjects. But our data shows that they did use null subjects at this age. Furthermore, the English-speaking children were different from the Chinese-speaking children, in that the latter used both null subjects and null objects during the time we tested them. These observations provide counterevidence to the Jaeggli and Hyams proposal.

This study also shows that although it is important to have theory guide research in the field of language acquisition, it is likely that the data will show where the predictions of the theory are in error, or where the theory needs refinement. Even if the parameter theory generally holds, it still could be true that the process of resetting might be slower for some parameters than others; in other words, in some aspects of the use of null subjects, the restructuring can be gradual and take a longer time than was previously thought. The result of this study also suggests that the linguistic environment or linguistic input shapes the child's grammar from a very early stage, e.g., as seen in the early cross-language differences in use of both null subjects and null objects.

REFERENCES

Bloom, P. (1990). Subjectless sentences in Child language. *Linguistic Inquiry, 21*, 419-504.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Gerken, L. (1990). A metrical account of children's subjectless sentences. In J. Carter, R. Dechaine, B. Philip, & T. Sherer (Eds.), *Proceedings of NELS 20, Volume 1*, (pp. 121-134). Amherst, MA: GLSA.

Huang, C. T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Doctoral Dissertation, MIT.

Huang, C. T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*,15, 531-574.

Huang, C. T. J. (1989). Pro-drop in Chinese A Generalized Control Theory. In Jaeggli & Safir (Eds.), *The null subject parameter*, (pp. 185-214). Dordrecht: Kluwer Academic Publishers.

Hyams, N. (1983). The pro-drop parameter in Child Grammars. *WCCFL, 2*, (pp. 126-139).

Hyams, N. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Kluwer Academic Publishers.

Hyams, N. (1989). The null subject parameter in language acquisition. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter*, (pp. 215-238). Dordrecht: Kluwer Academic Publishers.

Hyams, N. (in press). A reanalysis of null subjects in child language. In J. Weissenborn, H. Goodluck, & T. Roeper (Eds.), *Theoretical issues in language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hyams, N., & Wexler, K. (1991). *On the grammatical basis of null subjects in child language*. Manuscript, UCLA and MIT.

Institute of Linguistics, Academy of Sciences. (1973). *Modern Chinese Dictionary*. Beijing, PRC: The Commercial Publishing House.

Jaeggli, O., & Hyams, O. (1988). Morphological uniformity and the setting of the null subject parameter. In J. Blevins & J. Carter, (Eds.), *Proceedings of NELS, 18, Volume* 1. Amherst, MA: Graduate Linguistic Student Association.

Jaeggli, O., & Safir, K. (1989). The null subject parameter and parametric theory. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter*, (pp. 1-44). Dordrecht: Kluwer Academic Publishers.

Lasnik, H. (1989). On certain substitutes for negative data. In R. Matthews & W. Demopoulos (Eds.), *Learnability and linguistic theory* (pp. 89-105). Dordrecht: Kluwer Academic Publishers.

Lillo-Martin, D. (1986). *Parameter setting: Evidence from use, acquisition, and breakdown in American Sign Language*. Doctoral dissertation, University of California at San Diego.

Lillo-Martin, D. (1987). Parameter setting in the acquisition of American Sign Language. Presented at the Boston University Conference on Language Development.

Lillo-Martin, D. (1991). *Universal Grammar and American Sign Language: Setting the Null Argument Parameters*. Dordrecht: Kluwer Academic Publishers.

Mazuka, R., Lust, B., Wakayama, T., & Snyder, W. (1986). Distinguishing effects of parameters in early syntax acquisition: A cross-linguistic study of Japanese and English. *PRCLD, 25*, 73-82.

McCloskey, J., & Hale, K. (1984). The syntax of person-number inflection in modern Irish. *Natural Language and Linguistic Theory, 1*, 487-533.

McDaniel, D., & McKee, C. (in press). Which children did they show obey strong crossover? In H. Goodluck & M. Rochemont (Eds.), *The psycholinguistics of island constraints*.

Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry, 17*, 501-557.

Roeper, T. W. (1986). How children acquire bound variables. In B. Lust (Ed.), *Studies in the acquisition of Anaphora Vol. I*, (pp. 191-200). Dordrecht: Kluwer Academic Publishers.

Roeper, T., Rooth, M., Mallis, L., & Akiyama, S. (1984). The problem of empty categories and bound variables in language acquisition. Manuscript, University of Massachusetts, Amherst.

Sarma, J. (1991). *The acquisition of wh-questions in English*, doctoral dissertation, University of Connecticut.

Thornton, R. (1990). *Adventures in long-distance moving*, doctoral dissertation, University of Connecticut.

Weissenborn, J. (in press). Null subjects in early grammars: Implications for parameter setting theory. In J. Weissenborn, H. Goodluck, & T. Roeper (Eds.), *Theoretical issues in language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

FOOTNOTES

Language Acquisition, 2(3), 221-254 (1992).

† Also University of Connecticut

†† Also Wesleyan University

††† Also Wellesley College

[1] The following abbreviations are used in the glosses:

[e]: null argument

ASP: Aspect

DE (footnote 7); NE (p.14); MA (p.20): Chinese particles which have no stress, and no meaning of their own when used in a statement

BA (p.23): a passivizing morpheme in Chinese

[2] Chinese examples not otherwise credited are provided by QW.

[3] The null subjects in (5a, b & c) can be interpreted or understood as "sky."

[4] In his (1989) paper, Huang amends this option in a way which also allows the matrix subject to be *pro*, by saying that an empty pronominal (*pro*) must be identified by the closest nominal element *if there is one*. We will continue to adopt the (1984) analysis, by which only embedded subjects can be *pro*.

[5] Roeper, Rooth, Mallis, and Akiyama make this suggestion for a completely different reason. They discuss an experiment in which children appear to violate strong crossover for a long period of time. They account for this finding with the hypothesis that children begin with *pro* but not variables as empty categories. However, there is new evidence which suggests that children do not actually violate strong crossover (see McDaniel & McKee, in press, Thornton, 1990), and that they do have variables.

[6] The experimenter, QW, is a native speaker of Mandarin from the People's Republic of China. She is also fluent in English.

[7] None of the Chinese children in MLU group 3.5 (2-year-olds) and 4.5 (3-year-olds) produced any sentences with embedded clauses. Only one of the 4-year-olds (YD) produced few sentences with embedded clauses. However, all three of his sentences with embedded clauses were produced with an overt subject, e.g.,

Tā xiǎng, *Lǎo láng chūi* bù dǎo zhè mùtóu fángzi de.
He thought, old wolf blow not down this wood house DE
'He thought that *the old wolf* could not blow down the wood house.'

[8] Statistical comparison between the use of null objects by the American children and the Chinese children was unnecessary given the big differences between the ranges of the percentages.

[9] An interesting comparison can be made with the acquisition of German. Weissenborn (in press) claims that adult German is like Chinese in allowing null arguments identified by discourse topics, but he says that the occurrence of null arguments in German is more restricted than in Chinese, according to pragmatic factors. As he points out, German-speaking children will then need to learn those pragmatic factors which allow for null arguments in German on the basis of more linguistic experience than that which allows the Discourse Oriented Parameter to be set. He indicates that the development of the correct use of null arguments in German takes some time.

[10] CC=Chinese Children; AC=American Children; AAC=Adjusted American Children; CA=Chinese Adults.

APPENDIX 1: CHILD SUBJECTS

Subject	Language	age	Sex	MLU	Subj.drop	Obj.drop	Adj.Subj.drop
ZY	Chinese	2;0	F	2.41	48.103	15.952	
AN	Chinese	2;3	M	3.60	62.144	21.335	
WW	Chinese	2;5	F	4.23	56.937	23.077	
HE	Chinese	3;1	F	4.44	58.669	24.159	
LX	Chinese	3;4	M	4.27	44.532	12.827	
ZZ	Chinese	3;5	F	4.52	33.750	27.143	
SK	Chinese	4;1	M	5.04	45.439	22.479	
ML	Chinese	4;3	M	4.83	40.756	29.365	
YD	Chinese	4;4	M	5.98	28.572	26.250	
AR	English	2;5	M	2.69	58.636	8.333	51.177
SR	English	2;8	F	4.10	27.922	9.091	17.388
DS	English	2;10	F	3.74	17.156	7.500	9.091
EL	English	3;6	F	4.58	11.395	3.125	3.949
ER	English	3;8	M	4.80	25.981	5.179	5.390
DR	English	3;9	F	4.65	14.063	0.000	4.087
SP	English	4;2	F	4.49	59.524	0.000	18.831
SM	English	4;4	M	3.84	45.834	0.000	4.167
PT	English	4;5	M	4.51	37.436	0.000	17.179

APPENDIX 2: RESULTS FROM ADULT SUBJECTS

Subject	Subj.-drop	Obj.-drop
BM	33.670	6.719
BX	39.136	22.028
ET	43.363	10.417
LM	32.834	10.976
LP	25.322	8.495
QG	26.423	7.143
QQ	40.94	11.334
WC	40.298	8.929
YL	43.177	6.667

APPENDIX 3: RESULTS FROM CHILD SUBJECTS

Mean percentages of sentences with null subjects and with null objects

Subj.[11]	Subj.-drop	(s.e.)	Obj.-drop	(s.e.)
CC	46.543	3.776	22.533	1.761
AC	33.105	6.120	3.572	1.313
AAC	14.584	5.025		
CA	36.129	2.296	8.387	2.123

Testing results arranged according to chronological age

Subj.	Age	MLU	Subj-drop	(s.e.)	Adj.SD	(s.e.)	Obj.-drop	(s.e.)
CC	2	3.41	55.728	4.098			20.192	2.165
CC	3	4.41	45.650	7.215			21.376	4.361
CC	4	5.28	38.252	5.026			26.031	1.991
AC	2	3.51	34.571	12.427	25.885	12.871	8.308	0.459
AC	3	4.65	17.146	4.484	4.475	0.459	2.948	1.653
AC	4	4.28	47.597	6.437	13.392	4.637	0	0

Testing results arranged according to MLU

Subj.	Age	MLU	Subj-drop	(s.e.)	Adj.SD	(s.e.)	Obj.-drop	(s.e.)
CC	2	3.41	55.728	4.098			20.192	2.165
CC	3	4.41	45.650	7.521			21.376	4.361
CC	4	5.28	38.252	5.026			26.031	1.991
AC	2	3.51	34.571	12.427	25.885	12.871	8.308	0.459
AC	3,4	4.48	32.372	7.660	8.933	2.884	1.474	0.991

APPENDIX 4: THE FOLLOW-UP STUDY

Subject	Total # of sentences	# of sentences with transitive verbs	% Subj.-drop	% Obj.-drop
HD	295	176	41.36	38.07
HH	264	132	47.73	43.94
LQ	288	97	38.54	35.05
SL	316	122	49.68	39.34
TJ	344	167	50.87	44.31

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 251-254

Amplitude as a Cue to Word-initial Consonant Length: Pattani Malay*

Arthur S. Abramson†

Word-initial Pattani Malay consonants are short or long. The closures of the "long" consonants are longer than those of the "short" ones; this is a sufficient cue for perception, but in voiceless plosives the duration of the silent closure is audible only after a vowel, yet listeners label such isolated words well and so must use other cues. The peak amplitudes for the first syllables of disyllabic words are greater for initial long plosives. In this study, increments of closure duration and amplitude were pitted against each other for original short plosives and decrements for original long plosives. In tests, duration was by far the more powerful cue, although amplitude did affect the category boundary. By itself, however, amplitude is a weak cue. Further work is planned on the possible role of the shaping of the amplitude contour.

1. INTRODUCTION

Many languages are described as having a phonological distinction of length in vowels or consonants, or even both. If the term is taken literally, we would expect to find that the underlying mechanism is control of the relative timing of the articulators. Even so, a single mechanism might have a number of acoustic consequences, each of which could help in perception.

Pattani Malay, spoken by about a million ethnic Malays in southern Thailand, is unusual not only in having a length distinction for consonants in word-initial position but also in having one that is relevant for all phonetic classes of consonants in that position (Chaiyanara, 1983). Here are some minimal pairs of words showing the contrast:

/labɔ/	'to profit'	/lːabɔ/	'spider'
/make/	'to eat'	/mːake/	'eaten'
/bule/	'moon'	/bːule/	'months'
/katoʔ/	'to strike'	/kːatoʔ/	'frog'

The work was supported by NICHD Grant HD-01994 to Haskins Laboratories. The fieldwork in Thailand was made possible by a sabbatical leave from The University of Connecticut in 1988. I am grateful to the National Research Council of Thailand, the Department of Islamic Studies of The Prince of Songkhla University, Pattani, and the Department of Linguistics of Chulalongkorn University, Bangkok for their warm hospitality and help.

If, indeed, the crucial aspect of the articulatory gesture is the duration of the closure or constriction, for pairs like the first two it would not surprise us to find that the length distinction is quite discernible whether in utterance-initial or intervocalic position. But what about the stop consonants, especially the voiceless unaspirated stops of the language? The voiced stops do have voicing lead, so if you are close enough, you can hear short or longer stretches of glottal pulsing during the occlusion. The occlusions of the voiceless stops, however, are silent.

In earlier work (Abramson, 1987), I presented acoustic measurements of closure durations for the language, showing that the putative length categories are well separated by duration. Of course, the voiceless stops could not be measured in utterance-initial position. In another study (Abramson, 1986) I demonstrated, by systematically increasing the durations of short closures and decreasing the durations of long closures, that this feature is a sufficient and powerful acoustic cue for the perception of the distinction.

As for the voiceless stops, it was conceivable that the two categories were auditorily distinguishable in medial position only. This turned out not to be so in my control tests with unaltered words. Doing only slightly worse than with the

other classes of consonants, native speakers rather accurately identified short and long voiceless stops in isolated words. Among the various plausible acoustic effects of the mechanism, the most likely for the largely disyllabic words involved, was the peak amplitude of the first syllable relative to the second. Indeed, measurements (Abramson, 1987) revealed that this ratio is greater for long plosives, that is, both stops and affricates. Presumably, greater air pressure accumulated behind the occlusion before release accounts for the differences. Although both voiced and voiceless plosives showed a significant difference, the level of significance was higher for the latter. No doubt, this is to be explained by differences in glottal impedance of the airflow. The difference is not significant for the continuants.

2. PROCEDURE

This paper is a progress report of my test of the hypothesis that the peak amplitude of the first syllable relative to the second in disyllabic words is a sufficient cue for the perception of the distinction between short and long voiceless stops in Pattani Malay. For my major experiments, as part of an interest in combinations of phonetic features underlying the same phonemic distinction, I have pitted variants in duration and amplitude against each other to determine their relative power.

2.1. Control tests

Although the identifiability of initial short and long consonants had been demonstrated (Abramson, 1986), it seemed desirable also to do control tests for the recordings of my new speaker for this study. For each of seven minimal pairs of words I prepared a test containing 20 tokens of each of the two words, yielding 40 randomized stimuli. There were two such randomizations for each word pair. The nasal, lateral, fricative, and plosive categories were represented. The plosives included voiced and voiceless stops and voiceless affricates. (Unfortunately, my only pair of voiced affricates included a word, as I learned later, that would have embarrassed the women among the subjects, so I could not use that test.) The subjects were 30 undergraduate students, all native speakers of Pattani Malay, at the Prince of Songkhla University, Pattani, Thailand.

2.2. Amplitude vs. duration

To test for the relative power of amplitude and duration, three pairs of words with velar, dental, and labial short and long stops respectively were used. All of them were recorded at the end of the carrier sentence /dio katɔ/ 'he said.' By means of the Haskins Laboratories Waveform Editing and Display System (WENDY), the stop closure of the short member of each pair was lengthened in 20-ms steps until it reached or exceeded the duration of its long counterpart. The closure of the long member was shortened in the same way. The first syllable of each variant of the original short stop was increased in amplitude in five 2-dB steps. Likewise, the first syllable of each variant of the original long stop was *decreased* in amplitude in five 2-dB steps. Two test orders were recorded from randomizations of two tokens each of all the resulting stimuli and played to 30 native speakers for identification of the key words.

2.3. Amplitude in isolated words

The perceptual efficacy of amplitude without help from closure duration was tested by taking all the amplitude variants from the original short and long forms of one of the word pairs in section 2.2. Two test orders were recorded from randomizations of four tokens of each stimulus and played to 30 native speakers.

3. RESULTS

3.1. Control tests

The previously demonstrated identifiability of the utterance-initial consonants (Abramson, 1986) was reaffirmed. The major difference is that the voiceless long affricates in this sample were labeled correctly 96% of the time, whereas in the last study it was just above chance at 55%.

3.2. Amplitude vs. duration

Because of the limitation on space, the results of only two of the experiments are given here. Figure 1 gives the responses of 30 native speakers to nine durations in 20-msec steps of the [k]-closure in /kameŋ/ 'goat' combined with six amplitude levels in 2-dB steps. The vertical axis shows the percentage identification as short /k/. The earlier crossover of the higher-amplitude curves at the 50% point to the long-/k:/ category, giving judgments of /kːameŋ/ 'goatlike,' is highly significant [F (40, 1160)=9.0, $p < .001$]; nevertheless, the values of duration at the short end are very little affected. The opposite procedure, shortening original long /kː/ and lowering the amplitude, yielded similar results, as shown in Figure 2. The results are essentially the same for the other two places of articulation.

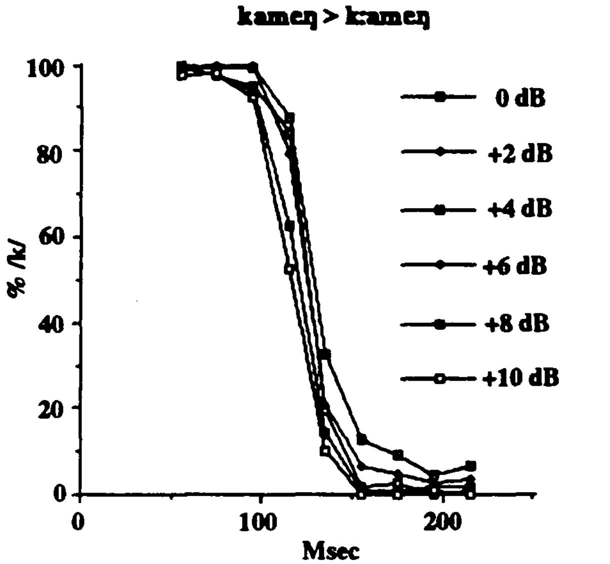

Figure 1. **Responses to /kameŋ/ 'goat' and its variants with increased closure duration and first-syllable amplitude.**

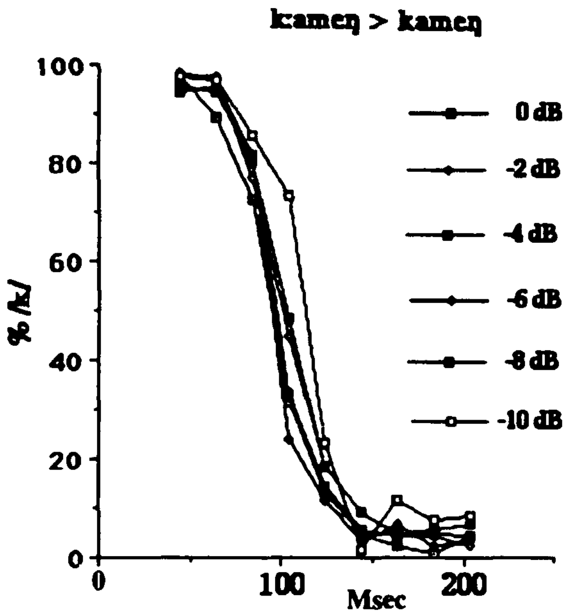

Figure 2. **Responses to /kːameŋ/ 'goatlike' and its variants with decreased closure duration and first-syllable amplitude.**

3.3. Amplitude in isolated words

In Figure 3 both the short and long responses are plotted for increments of amplitude on original /pagi/ 'morning.' While the two curves converge, they never cross each other. Figure 4 shows rather similar effects for decrements of amplitude combined with isolated tokens of /pːagi/ 'early morning.'

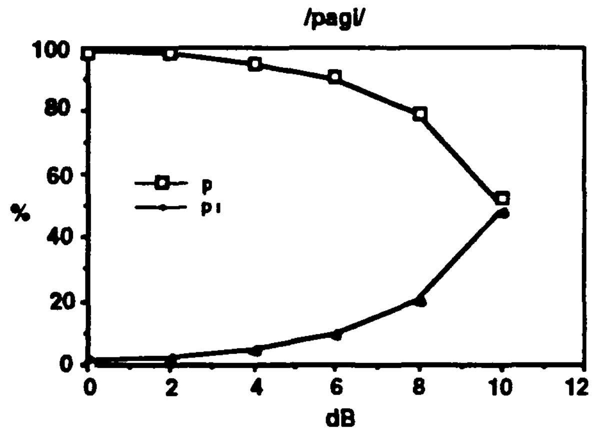

Figure 3. **Responses to isolated /pagi/ 'morning' and its variants with increased first-syllable amplitude.**

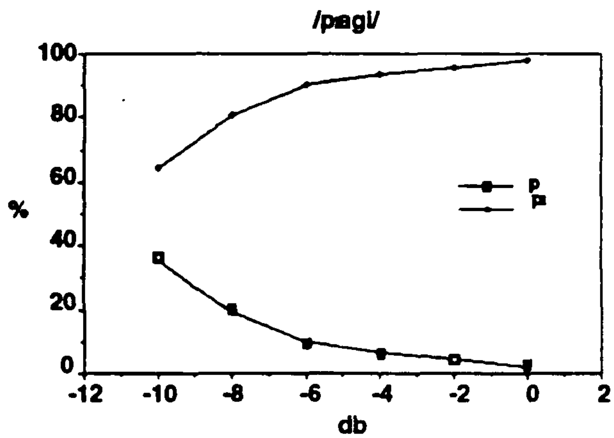

Figure 4. **Responses to isolated /pːagi/ 'early morning' and its variants with decreased first-syllable amplitude.**

4. CONCLUSION

It is clear that when both features are present, duration is dominant; nevertheless, the boundary between the two perceptual categories is significantly affected by relative amplitude. In utterance-initial position, however, relative amplitude is only a weak cue, apparently secondary to something else.

To understand how the length distinction is perceived in utterance-initial voiceless plosives, perhaps further work should be done on the possible role of the shaping of the amplitude contour. That is, maybe a finer analysis of utterances and a more complicated making of stimuli will show, for example, that the rise-time of the amplitude carries more weight than the peak value, or that the two work together. Indeed, a very preliminary look at this time suggests that the rise time is shorter in the production of the

long stops. Also, it is possible that the major amplitude difference is confined to the region of the release burst. Other features that have not seemed promising so far, such as fundamental frequency and rate of formant transitions, may have to be examined more closely too.

REFERENCES

Abramson, A. S. (1986). The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association, 16*, 8–16.

Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. In *Proceedings of the XIth International Congress of Phonetic Sciences., 6* (pp. 66–70). Tallinn: Academy of Sciences of the Estonian S.S.R.

Chaiyanara, P. M. (1983). *Dialek Melayu Pattani den bahasa Malaysia*. Kuala Lumpur: Master's thesis, University of Malaya.

FOOTNOTES

*In *Proceedings of the XIIth International Congress of Phonetic Sciences, Vol. 3* (pp. 98-101). Aix-en Provence: Université de Provence, France.

†Also the University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 255-262

Tone Splits and Voicing Shifts in Thai: Phonetic Plausibility*

Arthur S. Abramson† and Donna M. Erickson††

At the time of the emergence of its daughter languages, Proto-Tai is said to have had three phonemic tones on "smooth" syllables and four voicing categories for initial consonants, which would have been inherited by Old Thai (Siamese). Correlations between tones and initial consonants across the Tai languages have led to the positing of tonal splits conditioned by the voicing states of initial consonants with a subsequent shifting of voicing features in certain lexical classes. This change purportedly underlies the system of five tones and three consonantal voicing categories of modern Thai. Thus for each tone of Old Thai, words with initial voiced consonants developed a lower tone and words with initial voiceless consonants, a higher tone.

It has been shown for a number of languages that right after the release of a voiced stop consonant the fundamental frequency (F_0) of the voice is likely to be lower than after the release of a voiceless stop and that such F_0 perturbations can influence phonemic judgments of voicing. This led to the designing of two experiments to test the phonetic plausibility of the argument: (1) CV syllables were synthesized with three values of voice onset time (VOT) acceptable as Thai /b p ph/. Each of these was combined with a continuum of F_0 contours that had previously been divided perceptually into the high, mid and low tones. These syllables were played to native speakers of Thai for tonal identification. (2) Labial stops with nine values of VOT separable into /b p ph/ categories were coupled on synthetic mid-tone and low-tone CV syllables with upward and downward F_0 onsets varying in extent and duration. The resulting syllables were played for identification of the initial consonants. The historical argument receives modest support, especially from the second experiment, suggesting that during a period of tone splitting, under the influence of audible F_0 perturbations, speakers could have brought about the rephonemicization of the old consonant categories. Thus, these results give direct support to the argument that pitch factors led to voicing shifts but only indirect support to the claim that they gave rise to tone splits.

INTRODUCTION

If the distinctive tones of present-day Central Thai (Siamese) are the outcome of a series of developments over the centuries from an early simpler Proto-Tai tone system, or even a pristine state of tonelessness, we are beset with a problem common to all diachronic phonology. Can the causes of sound change be found? For Thai, as for some other Asian languages, this problem is complicated and made even more interesting by an apparent intersection of changing tonal features and shifting voicing states of word-initial consonants. It is our wish here to try to shed phonetic light on this aspect of the history of Thai.

In learning their language, children are likely to deviate ever so slightly in pronunciation habits from their adult models in ways that are largely unnoticeable at the time (Gray, 1939; Vendryes, 1923). Insofar as these shifts are not random, they may accumulate gradually over the generations,

The first author's current work on this study has been supported by NIH Grant HD-01994 to Haskins Laboratories, that of the second author by a contract between the Advanced Telecommunications Research Institute International and Osamu Fujimura of Ohio State University. The first author also thanks the Ford Foundation for its research grant to him for the 12 months of the 1973-74 academic year. He is also grateful to the National Research Council of Thailand, Ramkhamhaeng University, and the now defunct Central Institute of English Language for their hospitality.

resulting in sound changes with phonological consequences. Linguists have concentrated on these structural alterations, describing them systematically and purporting to show that, by and large, they are so regular that they can be stated in terms of "laws" for individual languages or language families. Except for noting that most of these changes, once they have been traced, are not phonetically improbable—e.g., /m/ is not likely to become /g/—they seldom find underlying phonetic mechanisms that might have brought these changes about.

With the recent advance of our understanding of the production and perception of speech, it is tempting for the experimental phonetician to believe that phonetic hypotheses on the causes of sound change should be testable in the laboratory (Ohala, 1974). For such research, without any way to resurrect long-dead informants for a brief stint of field work, the most that we can hope to do is to test the phonetic plausibility of these hypotheses by using present-day speakers. It must be stressed that it is only the plausibility of a posited causal relationship between sound change and particular phonetic mechanisms that can be tested.

A number of studies on the plausibility of postulated phonetic mechanisms of change have appeared in recent years. For example, Whalen and Beddor (1989) have published experimental data compatible with an explanation of the rise of a nasal feature in Eastern Algonquian. As for the emergence of distinctive tones, Hombert, Ohala and Ewan (1979) have provided an excellent critical review of the instrumental and experimental work on this topic.

The term *tonogenesis*, apparently first used by James Matisoff (1970, 1973), can mean the emergence of phonologically distinctive tones in a previously toneless language under the influence of certain contextual features. Another use of the term has been as a label for the splitting of old tonal categories into a larger number of tones. J. Marvin Brown (1975) speaks of the "great tone split...that swept through China and northern Southeast Asia nearly a thousand years ago."

During the time of the emergence of its daughter languages, Proto-Tai is generally said to have had four voicing categories for initial consonants and three phonemic tones on "smooth" syllables, i.e., those ending in a nasal, glide, or long vowel, which would all have been inherited by Old Thai (Siamese). If we make our focus for the moment not the tones but the initial consonants, we find the consensus of the various sources (e.g., Li, 1977) to be that the voicing states of some of these consonants changed under the influence of the pitch slopes as the tones emerged. We epitomize the situation with the labial stops:

Proto-Tai	*ʔb	*b	*p	*ph
Central Thai	b	ph	p	ph

We see that in modern Central Thai we have /ph/ from two sources, as is reflected in the Thai writing system. The correspondences are not exactly the same for all Tai varieties; for example, in Chiangmai /*b/ > /p/. Our emphasis here, however, is on Central Thai. The phonetic nature of /*ʔb/ is problematic (see Erickson, 1975 for a discussion). Haudricourt (1956) makes the rather tempting suggestion of [b^h] as an intermediate stage in the shift from /*b/ to /ph/.

With help from the writing systems, study of correlations between tones and initial consonants has led to the positing of tonal splits conditioned by the shifting voicing states of those consonants (Haudricourt, 1956; Li, 1947, 1977; Maspero, 1911). That is, ignoring the special problem of one of the four classes of consonants, the so-called glottalized consonants (see Erickson, 1975), we find that for each tonal category of Old Thai words with initial voiced consonants developed a lower tone and words with initial voiceless consonants, a higher tone. Thus the three Proto-Tai tones on smooth syllables, named simply A, B. and C[1] in the absence of knowledge of their phonetic nature, would have split into six. In fact, given the vicissitudes of the spread of phonological change over related languages, we find that Central Thai, which is the dialect of the Bangkok region and the basis of the official language of Thailand, has only five tones, while other regional dialects and other Tai languages have six or more, with differences among them in pitch contours as well. In a chart adapted from the work of Fang Kuei Li (1977, pp. 24–33), we give an outline of the tonal shifts from Proto-Tai to Central Thai on page 257.

Aside from these historical hypotheses, it has been known for some time that in human speech the fundamental frequency (F0)[2] of a syllable beginning with a voiced consonant is likely to be lower, for at least part of its duration, than that of a syllable beginning with a voiceless consonant (House & Fairbanks, 1953; Lehiste & Peterson, 1961). Indeed, it is remarkable that the early historical linguists logically inferred this likelihood without access to supporting physiological and acoustic phonetic research!

PROTO-TAI		CENTRAL THAI				
Tone	**Initial**	**Tone**	**Examples**			
A	Voiceless	Mid or rising	**paj**	to go	**fǒn**	rain
	Voiced	Mid	naa	rice field	wan	day
B	Voiceless	Low	**kàw**	old	**phàa**	to split
	Voiced	Falling	**phɔ̂ɔ**	father	**nâŋ**	to sit
C	Voiceless	Falling	**kâw**	nine	**nâa**	face
	Voiced	High	**thɔ́ɔŋ**	belly	**máa**	horse

For Thai (Erickson, 1975; Gandour, 1974) and other languages (Hombert, 1975), it has been found that F0 is likely to rise upon release of a voiced initial and fall upon release of a voiceless initial; both of these perturbations tend to end and blend in with the prosodic pattern of the syllable as determined by the sentence intonation and, in tone languages, the lexical tone. Other studies (e.g., Kohler, 1982; Löfqvist, Baer, McGarr, & Story, 1989; Ohde, 1984; Umeda, 1981) do not support a clearcut dichotomy between rising and falling perturbations. Rather, the F0 upon release of the voiced stop may in fact be on a level with, or at least not separable from, the rest of the contour; it may even fall a bit, or it may indeed rise; the crucial difference is that it is lower than the F0 onset upon release of a voiceless stop.

Physiological basis. As shown in literature reviews (Erickson, 1975; Ohala, 1978; Hombert et al., 1979), much ink has been spilled in support of various mechanisms that might underlie the F0 differences. Varying amounts of air flow governed by glottal size do not last long enough after stop release to account for the full effect. The role of myoelastic factors has long seemed much more probable. This would have to be some kind of difference in tension of the vocal folds. The problem has been to demonstrate this and tell what the mechanism is. One conjecture was vertical tension (Halle & Stevens, 1971), although it was hard to see how this might be executed, in spite of the finding of a higher position of the larynx for voiceless stops (Ewan & Krones, 1974). We are convinced by the recent work of Anders Löfqvist and his colleagues (Löfqvist et al., 1989; Löfqvist & McGowan, in press) that responsibility lies with varying degrees of contraction of the cricothyroid muscle used for control of vocal-fold tension to maintain or suppress vibration. Greater amounts of tension to help suppress voicing upon opening the glottis, combined with aerodynamic consequences, will cause higher F0 values in the speech signal.

Perception. The historical argument depends, of course, on the audibility of the F0 differences. Through psychoacoustic tests, Hombert (1975) showed that F0 movements of comparable magnitude are discriminable. It has also been found that either in somewhat exaggerated form (Haggard, Ambler, & Callow, 1970) or within more or less normal ranges (Abramson & Lisker, 1985; Fujimura, 1971; Kohler, 1985; Silverman, 1986; Whalen, Abramson, Lisker, & Mody, 1990) F0 perturbations can influence judgments of voicing in stops in such languages as English, Japanese, and German.

Goals of this study. If we assume these findings in production and perception to be universal and thus relevant to Southwestern Tai, the branch that gave rise to Thai, we might suppose that speakers of the language, already accustomed to the three-way tonal contrast of Proto-Tai, were psychologically receptive to the pitch fluctuations normally occurring with voicing distinctions.[3] We might suppose that attention was gradually shifted from the increasingly unstable voicing states of the initial consonants to the effects of the pitch perturbations on the following vowels. Increasing awareness of the perturbations could have led, through auditory feedback to production mechanisms, to enhancement of the effect by means of articulatory reinforcement and exaggeration of pitch differences. In this way, phonemicization of the pitch fluctuations came about, yielding an increase in tonal categories and helping to keep the old lexical classes apart, while the consonantal voicing categories decayed, shifted, and even coalesced.

To examine the plausibility of the foregoing historical arguments, we carried out experiments on the possible perceptual interaction between tones and initial stop consonants in the Thai language of today. That is, on the assumption of *diachronic* interaction between initial consonants and tones, we tested two hypotheses on speakers of modern Central Thai: (1) Perturbations of fundamental frequency should affect the perception of voicing distinctions in initial stop consonants. (2) The voicing states of initial stop consonants should affect the perception of tones. It must be understood that for both hypotheses we are not saying that the factors mentioned will be *primary* for the perception of these phonological distinctions. Rather, support for the hypotheses will be obtained if the boundaries between the perceptual categories are significantly affected. So as to have incremental control over the dimensions of interest to us, we followed the common practice of using synthetic speech.

EXPERIMENT I: VOICE ONSET TIME

An underlying assumption in these experiments, borne out by earlier work, is that the voiced, voiceless unaspirated, and voiceless aspirated stops of Thai lie along a dimension of voice onset time (VOT), namely, the temporal relation between the closing of the glottis for audible pulsing and the release of the occlusion of the initial stop (e.g., Lisker & Abramson, 1964; Abramson & Lisker, 1965; Abramson, 1989). For /b d/, voicing begins somewhat before the release, yielding "prevoicing" or "voicing lead," i.e., audible glottal pulsing during the occlusion. For /p t k/, voicing begins at the release or shortly thereafter. For /ph th kh/, voicing begins somewhat after the release; during the resulting "voicing lag," turbulent air coming through the open glottis excites the supraglottal vocal tract, yielding aspiration. These differences along the VOT dimension have not only been found in the acoustic signals but have also been shown to be perceptually relevant.

Procedure. In Experiment I we replicated the old work on the perceptual efficacy of VOT in Thai in order to establish a baseline for the testing of our two hypotheses. Using the Haskins Laboratories parallel-resonance synthesizer, we made as our basic pattern for all stimuli a set of formant[4] transitions appropriate to the labial place of articulation[5] followed by three steady-state formants appropriate to the Thai long vowel /aa/. We set the voice source of the synthesizer to produce 37 VOT variants, ranging from 150 ms before the stop release to 150 ms after the release. We did this in 10-msec steps except for the region around the release, where we used 5-msec steps from 10 ms before the release until 50 ms after the release. Thus,. stops with VOT before the release, i.e., voicing lead, simulated varying amounts of closure voicing. All the rest of the stimuli had a silent labial closure. All VOTs after the release had their upper two formants excited by a noise source and no excitation in the first formant for the period of voicing lag to simulate aspiration with an open glottis. We made a satisfactory mid tone by means of a level F0 at 120 Hz with, for naturalness in utterance-final position, a slight fall at the end (Abramson, 1962). We prepared eight tape-recorded randomizations of the synthetic stimuli with two tokens of each one in each of the test orders. Thus, each subject could have responded 16 times to each stimulus; however, depending on their availability, the listeners varied somewhat in how many tests they took. We played the tests through headphones to 48 native speakers of Central Thai at Ramkhamhaeng University and the now defunct Central Institute of English Language for identification in Thai script as /baa/ 'teacher,' /paa/ 'to throw,' or /phaa/ 'to lead.'

Results. The results of Experiment I are given in Figure 1. The ordinate shows the percentage of responses to each stimulus as one of the voicing states, which are indicated by the coded lines. The VOT values of the stimuli are arrayed at the bottom along the abscissa.

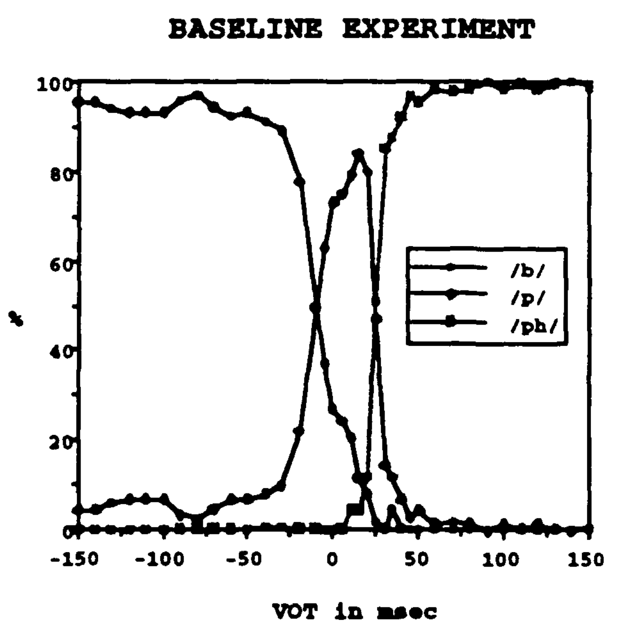

Figure 1. **Identification of synthetic labial stops varying in voice onset time. N=440.**

The category boundaries at the 50% crossover points, -7 ms for /b/–/p/ and 26 ms for /p/–/ph/, are very similar to those found in earlier work (Abramson & Lisker, 1965; Lisker & Abramson, 1970). Probably because of shortcomings in the synthesis, the /p/ category does not reach as high a peak as the other two. With these data in hand, we were ready to go on to Experiment II to test the first hypothesis.

EXPERIMENT II: F0 SHIFTS AND VOT

Procedure. We then turned to the matter of the effect of initial pitch perturbations on the identification of voicing states. We made our stimuli by varying the features of VOT and the extent of initial F0 shifts in the syllable pattern of Experiment I. With the data from Experiment I as a baseline, we chose nine VOT values to span the three voicing categories: -100, -20, 5, 10, 15, 20, 25, 30, and 80 ms. We imposed five F0 onsets upon each VOT variant. In addition to a flat onset at the 120 Hz level of our mid tone, we also had two downward shifts from 130 Hz and 140 Hz, as well as two upward shifts from 110 Hz and 100 Hz. Production data (Erickson, 1974) suggested that this 40-Hz range was reasonable. The shifts started at the first glottal pulse after the release of the stop and lasted 100 ms.[6] We presented three randomizations of the stimuli through headphones to 46 of our original listeners for identification as /b/, /p/, or /ph/ in Thai script, as in the previous experiment.

Results. The results of Experiment II are given in Figure 2. From top to bottom the three graphs show identification of the stimuli as /b/, /p/, and /ph/, respectively. Along the abscissa are displayed the VOT values, ranging from -100 ms to 80 ms. The ordinate shows the percentage of responses given to the various F0 conditions for each of the VOT values. There is a coded line for each of the F0 onsets.

An analysis of variance showed a high level of significance for the interaction between voicing state and F0 onset for /b/ and /p/: $F(8, 360)=2.67$, $p < .008$. Looking at the top graph, we see that the number of /b/ responses increases systematically as the initial F0 value decreases. That is, as the F0 value goes down, more stimuli are identified as /b/ at a later VOT value. In the middle graph, we again find an effect but in reverse.

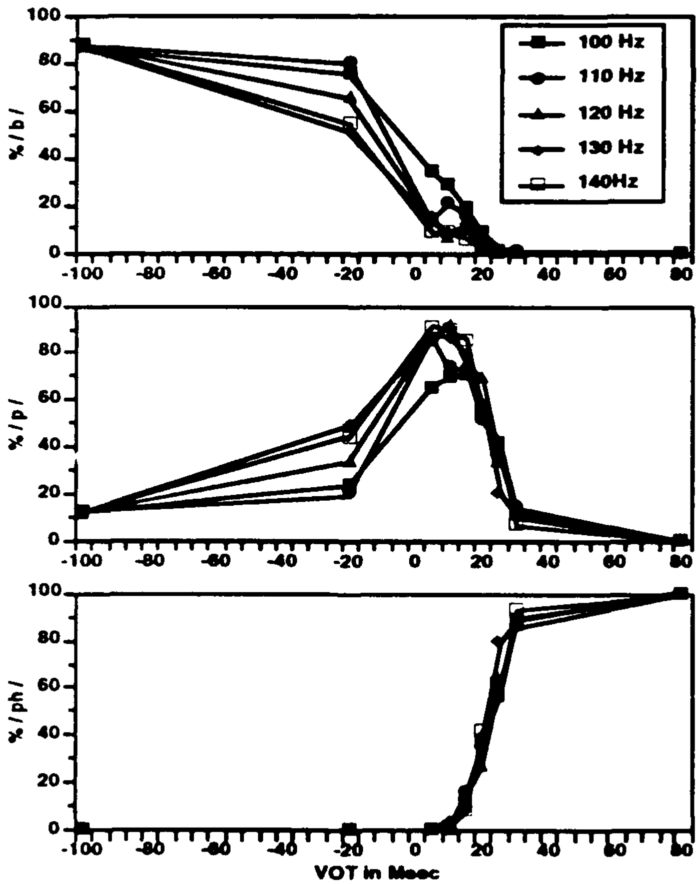

Figure 2. **Effects of F0 shifts on identifications. N=224.**

Higher Fo onsets increase the number of /p/ responses and thus yield earlier perceptual crossovers between /b/ and /p/. The bottom graph, however, shows a very tight clustering of the curves for /ph/ with no obvious effect of F0; this effect is not significant.

EXPERIMENT III: VOICING STATES AND TONE LABELS

We turned next to our second hypothesis, the one asserting that the voicing states of initial consonants will affect category boundaries for tones.

Procedure. To examine this question, we used a fan-shaped series of F0 contours with a common origin, which had previously (Abramson, 1978) been shown to be perceptually divisible into the three static tones, high, mid, and low. The 16 tonal variants all started at 120 Hz and moved to end points ranging from 152 to 92 Hz in 4-Hz steps. We synthesized syllables with VOT values suitable for /b p ph/ and formant frequencies for the vowel /aa/. The syllable meant to be heard as /baa/ had a VOT of -100 ms, /paa/, 0 ms, and /phaa/, 80 ms. The onset of each F0 contour began with the release of the stop. For /b/, the simulated closure, i.e., the 100 ms of voicing lead before the release, was at a level F0 of 100 Hz. Several randomized test orders were played through head phones to nine native speakers of Central Thai at The University of Massachusetts in Amherst.[7]

Results. The subjects fully accepted the three VOT values as the intended voicing states. Their tone labels are given in Figure 3. From top to bottom, the three graphs show how the listeners labeled the F0 contours as low, mid, and high tones respectively. The coded lines show the effects of the perceived voicing states of the stops on the tonal judgments. Along the abscissa are given the final F0 values of the tonal variants.

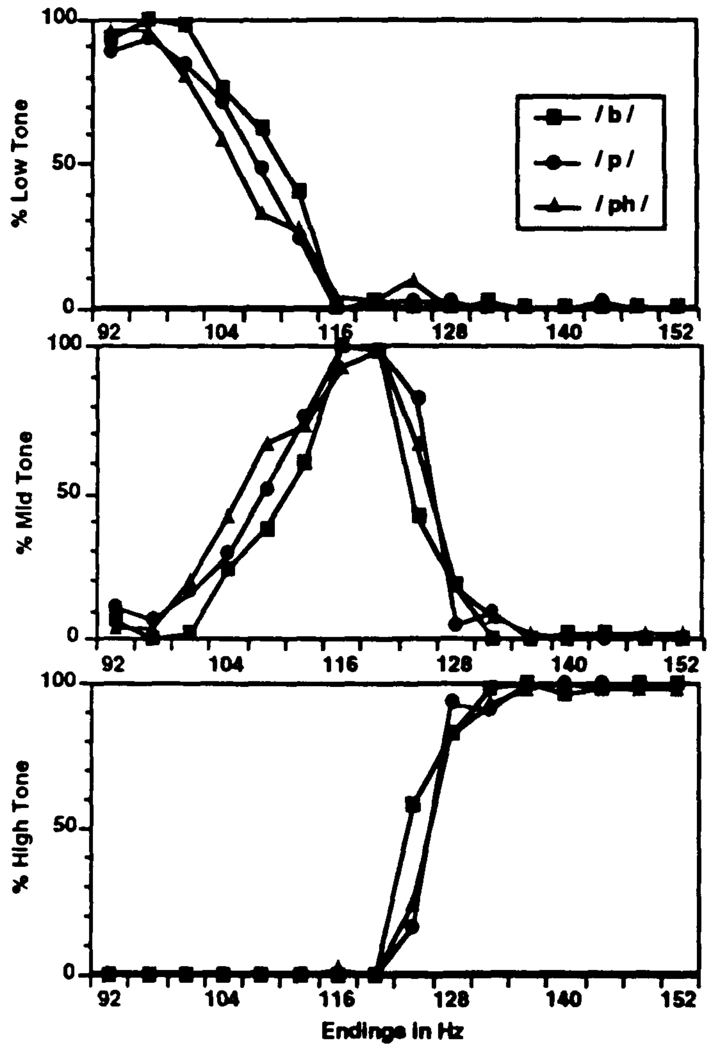

Figure 3. Effects of voicing states on tone labels.

Each point plotted gives the percentage of responses to that stimulus as the tone named on the ordinate.

An analysis of variance for the areas under the curves in the top graph shows a significant effect of the voicing states on the low-tone responses: $F(2, 16)=4.33$, $p < .04$. As shown by a post-hoc t-test, the main effect ($p < .05$) is that initial /b/ yielded a greater number of low-tone responses than the other two stops. We can see from the 50% crossover points that the final F0 can be higher for /baa/ than for /p ph/ and still be identified as a low tone.

The data plotted for the mid tone in the middle graph also show a significant interaction in an analysis of variance between voicing states and tone responses: $F(2, 16)=8.93$, $p < .003$. Post-hoc t-tests ($p <.01$) show that it is /phaa/ that has a larger number of mid-tone responses than the other two syllables.

As for the high tone in the bottom graph, again an analysis of variance shows a significant interaction: $F(2, 16)=8.23$, $p < .004$. Just as for the low tone, here too post-hoc t-tests ($p <.01$) show that the effect comes from the difference between /b/ and the other two stops. That is, initial /b/ gives a higher number of high-tone responses than /p/ or /ph/. Also, note the earlier 50% crossover point for /b/ between the mid and high tones. The crossover points for /p/ and /ph/, however, lie on top of each other.

CONCLUSION

It is clear from our data that fundamental-frequency perturbations can affect the placement of perceptual boundaries along the dimension of voice onset time. It is also true that the voicing states of initial stop consonants can affect the labeling of a continuous series of fundamental-frequency contours as tones. There are details of the various interactions, such as the seemingly paradoxically opposed boundary shifts for /baa/ with the low and high tones, that will require more thought and, perhaps, further investigation.

By and large, then, our perceptual data seem to support the historical arguments concerning interactions between tone splits and voicing shifts. As pitch perturbations loomed larger in the consciousness of the community and gradually took on a distinctive function, one might suppose that the voicing states of initial consonants would have been reassessed perceptually and rearticulated to furnish new production norms. A combination of these factors would have brought about shifts in tonal and consonantal categories.

REFERENCES

Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. Pub. 20. Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34*, 295—303.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21, 319-325.

Abramson, A. S. (1989). Laryngeal control in the plosives of Standard Thai. *Pasaa, 19*, 85-93.

Abramson, A. S., & Lisker, L. (1965) Voice onset time in stop consonants: acoustic analysis and synthesis. In D. E. Commins (Ed.), *Proceedings of the Fifth International Congress on Acoustics*. Ia (Paper A51). Liège.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Linguistic phonetics: Essays in honor of Peter Ladefoged* (pp. 295-303). New York: Academic.

Brown, J. M. (1975). The great tone split: Did it work in two opposite ways? In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 33-48). Bangkok: Central Institute of English Language.

Erickson, D. E. (1974). Fundamental frequency contours of the tones of Standard Thai. *Pasaa, 4*, 1-25.

Erickson, D. (1975). Phonetic implications for an historical account of tonogenesis in Thai. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Tai Linguistics in Honor of William J. Gedney* (pp. 100-111). Bangkok: Central Institute of English Language.

Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, 2, 327-335.

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jørgensen* (pp. 221-232). Copenhagen: Akademisk.

Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of Phonetics*, 2, 337-350.

Gray, L. H. (1939) *Foundations of language*. Macmillan.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 47, 613-617.

Halle, M., & Steven, K. (1971). A note on laryngeal features. *Quarterly Progress Report of the Research Laboratory of Electronics, MIT*, 101, 1989-213.

Haudricourt, A. G. (1956). De la restitution des initiales dans les langues monosyllabiques: Le problème du thai commun. *Bulletin de la Société de Linguistique*, 52, 307-322.

Hombert, J. (1975). *Towards a theory of tonogenesis: an empirical, physiologically, and perceptually-based account of the development of tonal contrasts in languages*. Unpublished doctoral dissertation, University of California, Berkeley.

Hombert, J., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language, 55*, 37-58.

House, A. S., & Fairbanks, C. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105-113.

Kohler, K. J. (1982). F_0 in the production of lenis and fortis plosives. *Phonetica, 39*, 199-218.

Kohler, K. J. (1985). F_0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America, 78*, 199-218.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, 33, 419-425.

Li, F. K. (1947). The hypothesis of a pre-glottalized series of consonants in primitive Tai. *Academia Sinica, 11*, 177-188.

Li, F. K. (1977). *A handbook of comparative Tai*. Honolulu: The University Press of Hawaii.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314-1321.

Löfqvist, A., & McGowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics, 20*, 93-110.

Maspero, H. (1911). Contribution à l'étude du système phonétique des langues thai. *Bulletin de l'École Française d'Extrême-Orient, 19*, 152-169.

Matisoff, J. (1970). Glottal dissimilation and the Lahu high-rising tone: A tonogenetic case-study. *Journal of the American Oriental Society, 90*, 13-44.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Consonant types and tones*. Southern California Occasional Papers in Linguistics, 1.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical Linguistics II: Theory and Description in Phonology* (pp. 353-389). North Holland.

Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*, 224-230.

Silverman, K. (1986). F_0 segmental cues depend on intonation: The case of the rise after stops. *Phonetica, 43*, 76-91.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *Journal of the Acoustical Society of America, 70*, 350-355.

Vendryes, J. (1925). *Language: A historic introduction to history*. A. Knopf (trans. by Paul Radin).

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica, 47*, 36-49.

Whalen, D. H., & Beddor, P. S. (1989). Connections between nasality and vowel duration and height: Elucidation of the Eastern Algonquian intrusive nasal. *Language, 65*, 457-486.

FOOTNOTES

*Appears in *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics* Vol. 1 (pp. 1-16). Bangkok: Chulalongkorn University.

†Also University of Connecticut, Storrs.

††Division of Speech and Hearing Science, Ohio State University, Columbus.

[1] Li also posits tone D on syllables ending in a stop consonant. There is no way of identifying it with any of the other tones. The phonological treatment of modern Thai, however, generally aligns the tones on such syllables with certain tones that occur on smooth syllables.

[2] The fundamental frequency of a complex sound wave is equivalent to the repetition rate of the vibrating source. Thus in speech the number of cycles of vibration per second of the vocal folds, given as a number of Hertz (Hz), is the fundamental frequency, which, of course, may vary continuously. It is the primary physical correlate of the sensation of pitch.

[3] J. Marvin Brown (1975, pp. 43–45) has offered some very interesting speculation on the matter.

[4] A formant is the acoustic consequence of a resonance of the vocal tract. An array of formants at different resonant frequencies will specify the spectrum of a vowel.

[5] As the vocal tract changes its shape through articulatory movement, the formants will necessarily shift in frequency. Such formant "transitions" furnish perceptual cues to place of articulation.

[6] Actually, we also used two other durations for the shifts, 50 and 150 ms. Inasmuch as we found no significant difference for the durations, we are presenting our data for the middle value only, 100 ms.

[7] Unfortunately, we had to prepare this experiment after we had both returned from Thailand, so we had to use a much smaller number of subjects; nevertheless, we obtained enough data for statistical treatment.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 263-276

A Constraint on the Expressive Timing of a Melodic Gesture: Evidence from Performance and Aesthetic Judgment*

Bruno H. Repp

Discussions of music performance often stress diversity and artistic freedom, yet there is general agreement that interpretation is not arbitrary and that there are standards that performances can be judged by. However, there have been few objective demonstrations of any extant constraints on music performance and judgment, particularly at the level of expressive microstructure. The present study illustrates such a constraint in one specific case: the expressive timing of a melodic gesture that occurs repeatedly in Robert Schumann's famous piano piece, "Träumerei." Tone onset timing measurements in 28 recorded performances by famous pianists suggest that the most common "temporal shape" of this (nominally isochronous) musical gesture is parabolic, and that individual variations can be described largely by varying a single degree of freedom of the parabolic timing function. The aesthetic validity of this apparent constraint on local performance timing was investigated in a perceptual experiment. Listeners judged a variety of timing patterns (original parabolic, shifted parabolic, and nonparabolic) imposed on the same melodic gesture, produced on an electronic piano under MIDI control. The original parabolic patterns received the highest ratings from musically trained listeners. (Musically untrained listeners were unable to give consistent judgments.) The results support the hypothesis that there are classes of optimal temporal shapes for melodic gestures in music performance, and that musically acculturated listeners know and expect these shapes. Being classes of shapes, they represent *flexible constraints* within which artistic freedom and individual preference can manifest themselves.

INTRODUCTION

Much has been written about music performance, with the emphasis generally being on the diversity among interpretations by different artists and in different historic periods. Yet, in each period (and quite likely across periods) there have also been generally accepted performance standards, which were reflected in music education, performance practice, and music criticism. The nature of these standards has been discussed in a number of treatises (most notably Lussy, 1882), but rarely in objective and quantitative terms. This is particularly true with regard to the expressive microstructure of performance—all those variations that are not easily captured in music notation but that are essential to the communicative function of interpretation. Musicians are usually only dimly aware of these variations, which they control intuitively rather than deliberately. Similarly, musical listeners perceive the structure and expression conveyed by these variations without being aware of the microstructure as such. It has been up to experimental psychologists to discover and measure these variations objectively (e.g., Palmer, 1989; Repp, 1990; Gabrielsson, Bengtsson, & Gabrielsson, 1983; Seashore, 1938/67; Shaffer, 1981).

This research was made possible through the generosity of Haskins Laboratories (Michael Studdert-Kennedy, president). Additional support came from NIH BRSG Grant RR-05596 to the Laboratories. A short version of this paper was presented at the Second International Conference on Music Perception and Cognition in Los Angeles, February 1992.

I am grateful to Pat Shove for many stimulating discussions.

Even though a number of studies of expressive microstructure have been published, they have rarely provided evidence of constraints on performance parameters. The principal reason is that they usually were based on very small samples of performances, so no statements could be made about the generality of particular microstructural patterns. Hypotheses about the generality of such patterns, as instantiated for example in the performance rules of Friberg (1991) or in the hierarchical timing model of Todd (1985), remain to be validated on large performance data bases. Moreover, studies of music performance have rarely combined measurements with formal perceptual evaluations to confirm the aesthetic validity of the hypothesized or measured patterns.

The work of Johan Sundberg and his colleagues is a significant exception (see Sundberg, Friberg, & Frydén, 1991). A study by Sundberg and Verrillo (1980) had a purpose very similar to that of the present research. These authors were concerned with the temporal shape of the *ritardando*, the gradual slowing of tempo commonly observed in performance at the ends of most compositions. They asked whether there was an optimal time course for this slowing down which performers observed and listeners expected. They selected 24 recordings of rhythmically uniform music, mostly by J. S. Bach, and measured the onset intervals between successive tones, whose reciprocals they then plotted as local tempo decreasing over time. Sundberg and Verrillo found that the average function resulting from these measurements could be described in terms of two linear segments, the second steeper in slope than the first. They also conducted a perceptual test in which musically experienced listeners were presented with excerpts that exhibited various forms of *ritardando*, some corresponding to the observed average function and others having deviant temporal shapes of various kinds. The listeners tended to prefer the *ritardandi* corresponding to the original performances.

In a later discussion of the same data, Kronman and Sundberg (1987) abandoned the bilinear model and instead fitted the average data points with a single curve (a square-root function), which they claimed was similar to that observed when other rhythmic motor activities, such as locomotion, come to a smooth halt. (Specific references to relevant literature were not given.) This function thus may represent a rather general constraint on the optimal shape of the musical *ritardando*.

Although these studies exhibit some methodological weaknesses[1] and therefore can only be regarded as preliminary, they nevertheless set a good precedent for the kind of approach to be taken in investigations of performance constraints.

The present investigation concerns possible constraints on the temporal shape of an expressive melodic gesture. (The performance-oriented term "melodic gesture" is used here to refer to a brief sequence of melody tones that is executed as a single expressive unit. The equivalent term "(rhythmic) group" is often used in the musicological literature.) By a constraint is meant a restriction on the performance patterns that occur in expert interpretations and that are judged acceptable by musically experienced listeners. Melodic gestures occur throughout Western music in most styles, and they come in a large variety of forms. It seems unlikely that all these forms are subject to any single performance constraint. The nature of these constraints may vary as a function of many factors, including tempo, metric and harmonic structure, style, and so on. Rather than searching for a universal constraint, the present study focused on the timing pattern of one particular melodic gesture. If it could be demonstrated that this pattern is subject to a significant constraint in performance and in perceptual judgment, this would at least provide an existence proof of such constraints on expressive microstructure. Moreover, by focusing on a specific case, the constraint can be characterized rather precisely. Questions about its origin and generality may then form the basis for future research.

The melodic gesture under investigation occurs in Robert Schumann's famous piano piece, "Träumerei" (No. 7 of "Kinderszenen," op. 15), whose score is shown in Figure 1. The melodic gesture moves from bar 1 into bar 2. (See Figure 3 below.) In its notated form, it consists of five eighth-notes ascending in pitch and a final longer note which repeats the pitch of the preceding eighth-note. The gesture recurs six times (eight times, if the obligatory repeat of the first eight bars is counted) during the piece, with some variations in key and interval structure. These recurrences are aligned vertically in Figure 1. The gesture is of central importance to the expressive quality of a performance of "Träumerei" and may be assumed to be given close attention by both performers and listeners.

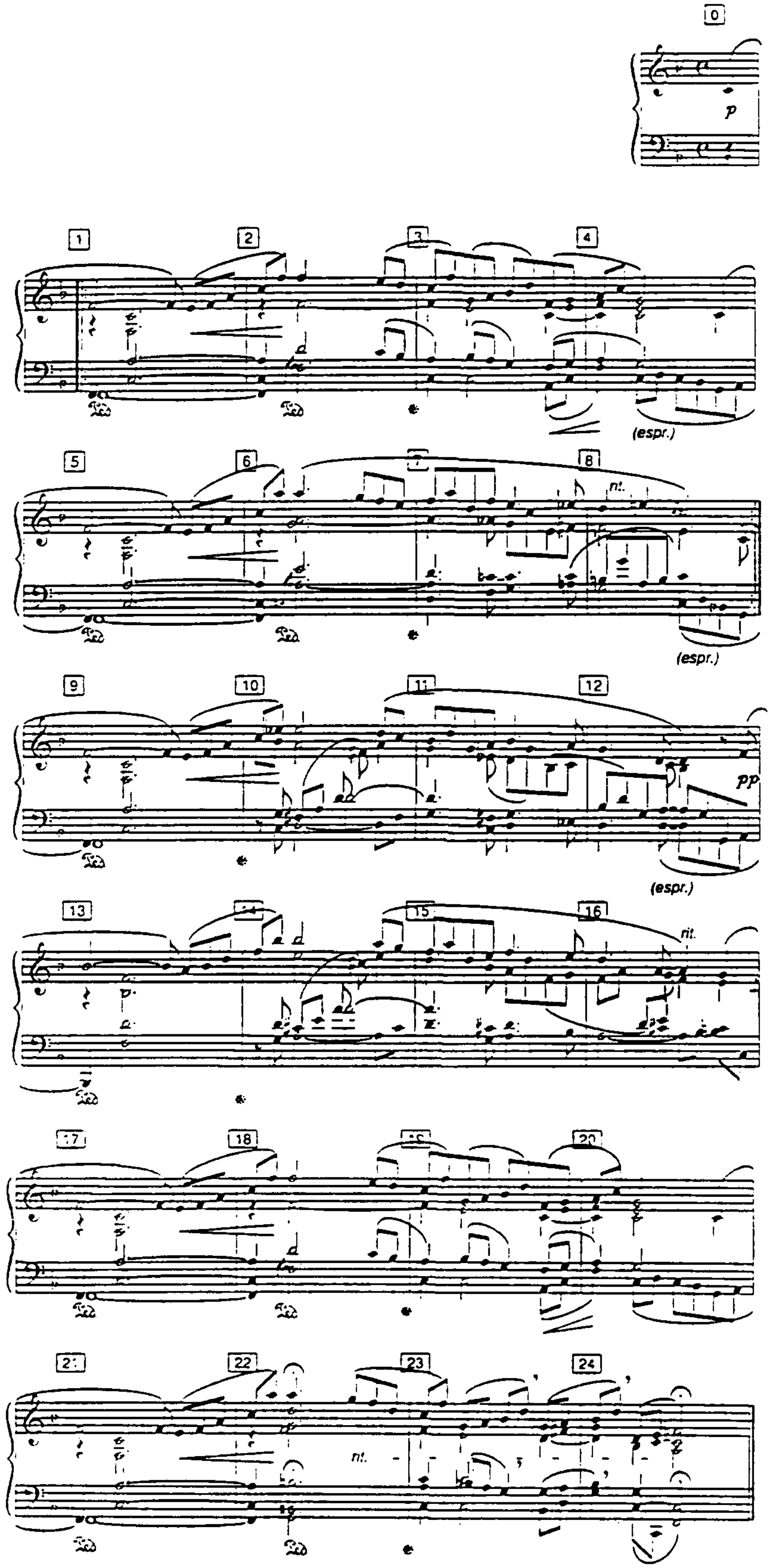

Figure 1. **Piano score of Schumann's "Träumerei," arranged on the page so parallel structures are vertically aligned. The score was created after the Clara Schumann edition (Breitkopf & Härtel) using MusicProse software; minor deviations from the original are due to software limitations.**

The onsets of the tones corresponding to the six melody notes define five interonset intervals (IOIs) which would be equally long if the music were performed mechanically (e.g., by a computer). In fact, they are never equal in a human performance; pianists always give an expressive temporal shape to this crucial part of the melody. This temporal shape can be visualized as the pattern of observed IOI durations, plotted as connected points equidistant along the x-axis ("score time"). How many such patterns are there? In principle, the melodic gesture can be performed with any temporal pattern whatsoever.[2] However, the hypothesis pursued here is that only certain patterns actually occur in expert performances and are found acceptable by listeners.

One characteristic of this class of patterns may be predicted on the basis of the general principle of final lengthening (e.g., Lindblom, 1978; Todd, 1985): A slowing down of tempo is often observed at the ends of action units such as phonological phrases in speech or melodic gestures in music, particularly when they coincide with the end of a larger structural unit, such as a clause (subphrase) or phrase. Therefore, the timing patterns to be investigated may be expected to show some lengthening of the last IOI(s). An independent reason for lengthening of the last IOI might be the occurrence of two grace notes (essentially a written-out *arpeggio*) in the left hand during that interval (see Figure 1). However, these grace notes occur only in bars 2, 6, and 18, not in bars 10, 14, and 22. The pattern of execution of these two sets of variants may differ. Another relevant phenomenon is the possible lengthening of accented tones. In the score, the fourth note of the melodic gesture follows a bar line and thus in theory carries a strong metrical accent (downbeat). Based on the notated music, therefore, a lengthening of the fourth intertone interval might be predicted. Musical intuition suggests, however, that this theoretical accent is suspended in performance, and that the accented tone of the melodic gesture is in fact the final one. Whether this is in fact so is an empirical issue to be addressed below. In principle, nothing can prevent a pianist from placing an overt accent on the fourth tone.

In the remainder of this paper, a summary of performance measurements is followed by the detailed report of a perceptual experiment. The measurements derive from a comprehensive analysis of timing microstructure in performances of Schumann's "Träumerei"; for details, the reader is referred to Repp (1992).

PERFORMANCE MEASUREMENTS

Tone onset timing measurements were obtained from the digitized waveforms of 28 different performances of "Träumerei," taken from commercial recordings (LP, CD, or cassette) by 24 pianists. Two famous pianists (Alfred Cortot and Vladimir Horowitz) were represented with three different recordings each. The measurements were averaged over the obligatory repeat of bars 1-8 (observed by all but two pianists in the sample) before further analysis. Thus there were data for six instances of the melodic gesture of interest in each of the 28 performances, a total of 168.

Initially, the geometric mean durations of the five IOIs for each of the six instances of the gesture were computed across the 28 performances. These durations (in ms) were plotted as a function of score time (i.e., at equal abscissa intervals), and their pattern was examined as to whether it could be fit by some simple function. These data are shown in Figure 2. It is evident that the timing pattern of each instance was fit well by a smooth curvilinear function, in fact a parabola (quadratic curve). Overall, pianists tended to speed up somewhat in the initial part of the melodic gesture and to slow down at the end. This slowing down was especially pronounced in the last instance of the melodic gesture (bars 21-22), where the score indicates a *fermata* (hold) on the last note. It was least pronounced in the two instances in the middle section of the piece (bars 9-10 and 13-14). All instances, however, were described well by quadratic functions which differed mainly in curvature.

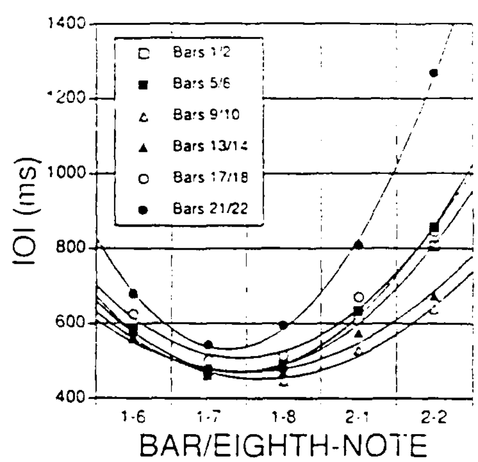

Figure 2. **Timing patterns of six instances of the same melodic gesture in "Träumerei." The data points are the geometric average durations of 28 performances (Repp, 1982), with quadratic functions fitted to them. The abscissa labels refer to bars 1-2.**

Subsequently, all 168 individual timing patterns were plotted and examined in the same way. It was found that 87% of them could be described rather well by quadratic functions of varying elevation (i.e., average tempo) and curvature (i.e., degree of tempo modulation). All but two of the exceptions followed a single pattern: a relative shortening of the last IOI.[3] This pattern, whose main representative was the French pianist Alfred Cortot in his three performances, suggests a different structural interpretation of the melodic gesture: a division into two subgestures and/or an intention to place an accent on the fourth tone. Three other pianists showed this pattern intermittently; Cortot himself consistently avoided it in bars 21-22, where he showed the standard parabolic timing curve.

Further analysis of the coefficients of the quadratic polynomials ($y = a + bx + cx^2$) fit to 87% of all instances revealed some strong relationships among the constant (a), linear (b), and quadratic (c) terms of these functions. The latter two, in particular, were highly correlated. There was also a substantial correlation between the quadratic and constant terms. Linear regressions among the coefficients made it possible to predict the linear and constant terms from the quadratic term and thus to generate a single family of parabolas by varying the quadratic term alone. (This family is shown in the upper left-hand panel of Figure 4 below.) It captures a substantial amount of the variance in the data, with deviations occurring mainly in the constant term (i.e., elevation along the ordinate, corresponding to variations in overall tempo), which is irrelevant to the temporal shaping of the melodic gesture.

These quadratic curves represent a strong constraint on the timing pattern of the melodic gesture studied here. Apparently, the large majority of expert pianists achieve a parabolic timing function by controlling a single degree of freedom. No pianist lengthened the second IOI, say, or shortened the first, or showed any pattern (other than the type favored by Cortot) that deviated substantially from a parabolic trajectory (though see Footnote 2). Even the Cortot pattern followed a parabolic curve through the first four IOIs. To the author, however, these performances sound mannered. This subjective impression, in conjunction with the overall predominance of parabolic timing patterns, suggested that the more typical parabolic patterns might also be preferred by other musically experienced listeners. This hypothesis was tested in the following perceptual experiment.

PERCEPTUAL EXPERIMENT

The purpose of this experiment was to demonstrate that listeners' aesthetic preferences converge on the timing patterns that characterize the majority of expert performances. To that end, subjects were presented with the melodic gesture of interest, executed with a variety of timing patterns, each of which was to be rated for acceptability on a 10-point scale. The timing patterns included parabolic and "hybrid" (nonparabolic) shapes. Among the former, there were some that belonged to the family of functions observed in actual performances, whereas others deviated in the location of the minimum. It was expected that listeners would prefer the "normal" over the deviant parabolic shapes. In addition, these functions varied in curvature. Since listeners might also exhibit a preference for a particular curvature (degree of tempo modulation) within each class of temporal shapes, some deviant shapes might actually be preferred over some normal shapes. However, for a given degree of curvature, the normal shapes were expected to be preferred most. The hybrid shapes were generated from two normal parabolic patterns of different curvature by interchanging their IOIs in all possible ways. It was expected that listeners' judgments would reflect the hybrids' degree of approximation to a parabolic shape. The responses were also expected to yield information about what deviations from the normal shapes are more readily tolerated than others. In fact, one of these deviant shapes resembled the Cortot type of pattern.

The role of listeners' musical experience was a rather crucial issue. If the parabolic constraint uncovered in the performance timing measurements reflects a general principle of physical motion, i.e., an optimal pattern of acceleration-deceleration, then even listeners without much musical experience might show a preference for it. The alternative possibility is that listeners need to be attuned to temporal patterns in classical music performance to show reliable preferences in this task. To investigate this issue, subjects both with and without musical experience were tested. A second question, concerning subjects with musical experience, was whether their judgments would be based on general knowledge of performance principles in classical music, or on specific knowledge of "Träumerei" and its performance. This question was not addressed rigorously, but some relevant information was obtained.

Methods

Subjects. Twenty-six subjects participated. The majority of them had responded to an advertisement in the Yale campus newspaper; others were recruited personally by the author and included some friends and family members who served without pay. Twelve subjects had little musical education; most of them did not play any instrument, while some had studied an instrument for a short time. Fourteen subjects were musically experienced; they included 11 pianists, two violinists, and one flutist, ranging in skill from advanced amateur to professional level.

Stimuli. The stimuli were generated on a Roland RD250S digital piano under MIDI control. Temporal resolution was 5 ms. Each stimulus consisted of the excerpt shown in Figure 3 (from bars 1-2 of "Träumerei"), played with one of 45 different timing patterns. The timing pattern was applied only to the melodic gesture of interest, which comprised five IOIs; the timing of the preceding context, comprising three longer IOIs, was constant at values representing the geometric means of the 28 expert performances measured by Repp (1992): 1065, 1380, and 1825 ms, respectively. The timing of the left-hand grace-note tones during the last IOI of the critical gesture was such that the first tone started after one third of the IOI had elapsed and ended with the onset of the second tone, which started after one third of the remaining interval had elapsed. (This timing pattern was fairly common in the 28 performances examined.) To make room for the grace-note tones, the preceding chord, the tied-over quarter-notes of the preceding chord in bar 2 were realized as tied-over eighth-notes. Sustaining pedal was added as indicated in the score. The tones had a fixed expressive intensity pattern similar to that of one of the expert performances.

The timing patterns of the critical melodic gesture are illustrated in Figure 4. The upper left-hand panel shows the five "normal" patterns, which followed parabolic functions of varying curvature. Each parabola was generated by the equation, $IOI(ms) = C + Lx + Qx^2$, where x stands for the ordinal numbers of the IOIs (1,...,5). The quadratic term (Q) of the polynomial equation was set at values of 20, 40, 60, 80, and 100, which span the range of most empirically observed timing functions. The linear (L) and constant (C) terms of the parabolas were derived according to the empirically determined regression equations, $L = 35 - 5.5Q$ and $C = 388 + 7.8Q$ (see Repp, 1992). The resulting stimuli were named Q20, ..., Q100.

The lower panels in Figure 4 illustrate two sets of deviant parabolic curves. Each set varied in Q along the same values as the normal set, but the constant and linear terms differed. In the "left-shifted" set, C was decreased by 300 and L was increased by 100, whereas, in the "right-shifted" set, C was increased by 300 and L was decreased by 100. In each case, the change in one parameter was arbitrary, but the change in the other parameter was chosen so as to keep the average IOI duration equal to that of the normal condition with the same Q. The stimuli in the left-shifted set (Q20L, ..., Q100L) started faster and ended slower than the normal stimuli; the opposite was true for the stimuli in the right-shifted set (Q20R, ..., Q100R).

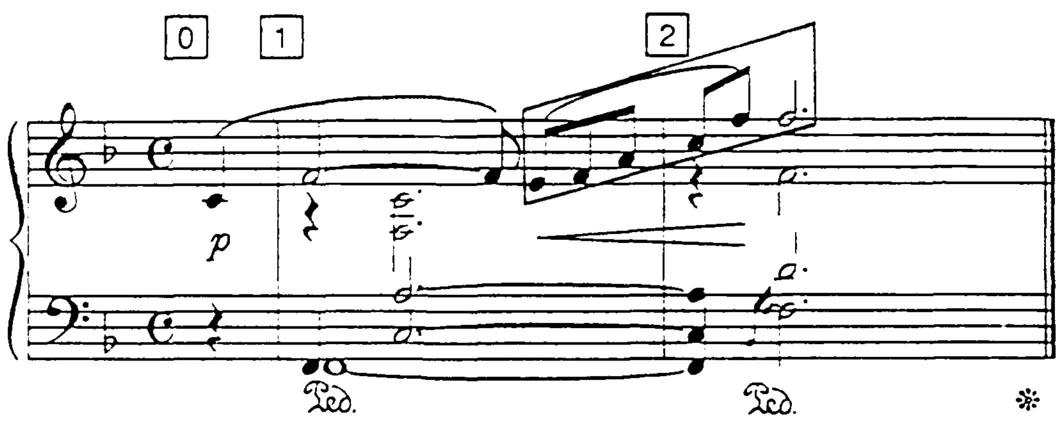

Figure 3. **The musical excerpt used in the experiment (from bars 1-2 of "Träumerei," with slightly modified final notes). The melodic gesture of interest is boxed in.**

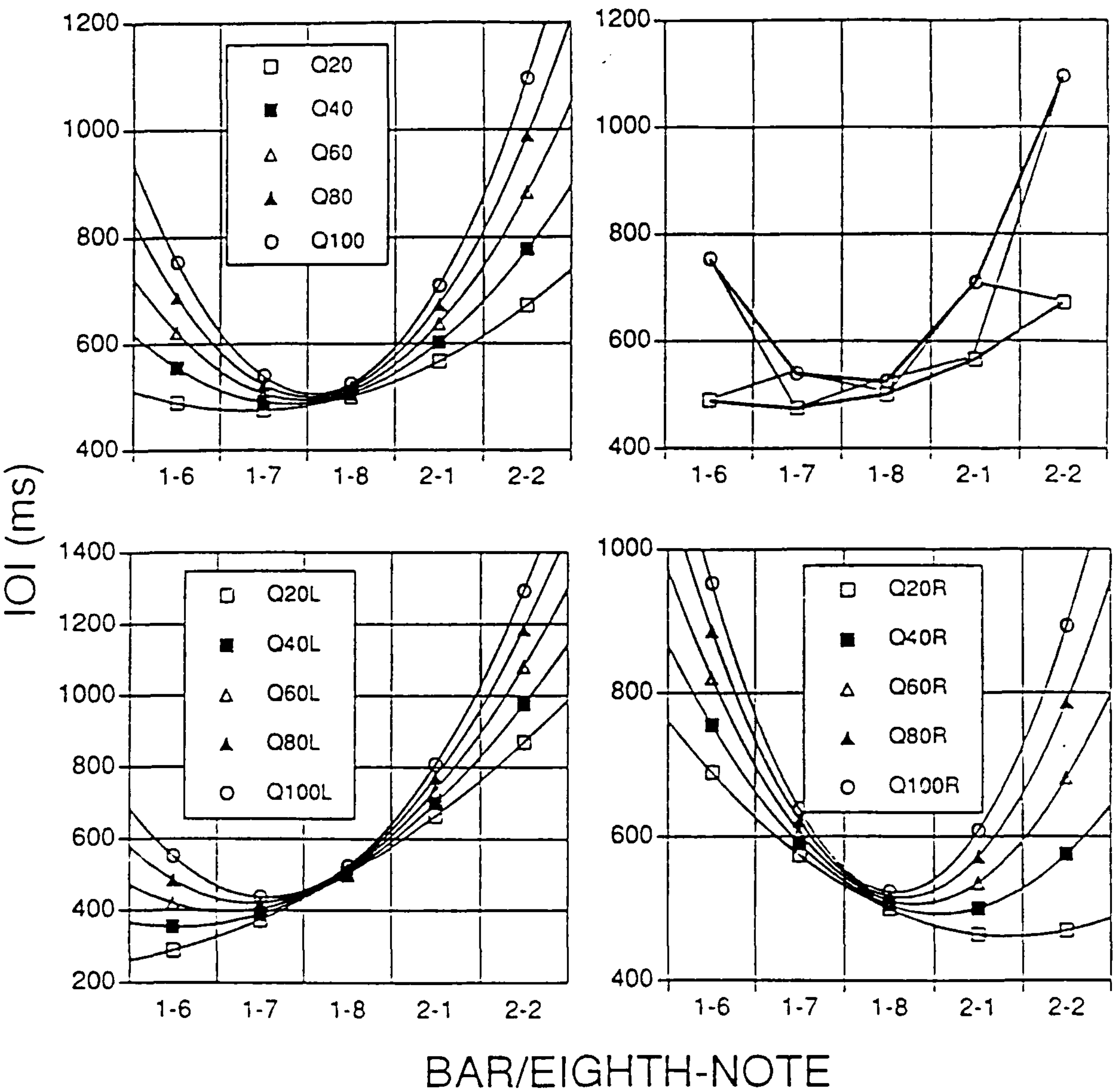

Figure 4. **Timing patterns of the experimental stimuli. Upper left-hand panel: normal parabolic patterns. Lower left-hand panel: left-shifted parabolic patterns. Lower right-hand panel: right-shifted parabolic patterns. Upper right-hand panel: hybrid patterns.**

The remaining 30 timing patterns were generated as illustrated in the upper right-hand panel of Figure 4. The heavy lines in the figure illustrate the Q20 and Q100 timing patterns, represented here as polygons rather than as smooth curves. Thirty hybrid patterns were generated by interchanging IOI durations from those two patterns. With two possible values for each of five IOIs, there are 32 possible patterns, two of which are the original ones. The original patterns were coded arbitrarily as H00000 (= Q20) and H11111 (= Q100), and hybrid patterns were coded as H10000, H11000, etc. Clearly, some of these hybrids (e.g., H00100, H11011) were very similar to the originals, whereas others were more dissimilar. Although some of them were clearly nonparabolic (e.g., H01010), others might by fit by a left-shifted or right-shifted parabola (H00011 and H11100, respectively). In contrast to the left- and right-shifted parabolic patterns, however, all individual IOIs in the hybrid patterns were within the normal range. One hybrid pattern, H11110, was not unlike the Cortot pattern described above.

The stimuli were recorded electronically from the audio output jack of the digital piano onto high-quality cassette tape . Six examples were recorded at the beginning of the tape, the first three with isochronous timing of the melodic gesture (i.e., with constant IOIs of 500 ms), and the second three being stimuli H01101, Q80R, and Q40. These examples were followed by three different randomized sequences of the 45 stimuli. Interstimulus intervals were 5 s, with an additional 5 s after each group of 15, and another 5 s between blocks.

Procedure. Subjects received a dubbed copy of the master cassette, accompanied by detailed printed instructions, an answer sheet, and a questionnaire about their musical experience. They listened on their home audio equipment and returned the completed materials. (Control over sound quality and playback level was not crucial in this study.)

The instructions displayed the score of he excerpt (cf. Figure 3) and included the following crucial sections:

> ...Each time the excerpt will be played with a slightly different timing pattern of the notes. Your task is to judge the aesthetic appeal of each timing pattern. Clearly, there are no right or wrong responses here; I want to find out what sounds good to *you*...."

(After the first set of examples had been introduced:)

> "...In the following three examples, the eighth-notes vary in duration, as they would in a human performance. Each of the three examples has a different timing pattern, and they may not (in fact, should not) sound equally good to you. Clearly, there are some timing patterns that are preferable to others. ... In the following test, you will indicate [your] preference by giving a numerical rating between 1 and 10 to each excerpt you hear, where 10 is the best possible rating and 1 is the worst. ... However, don't use these [ratings] in an absolute sense, but try to adjust to the diversity of timing patterns you hear and use the whole scale; that is, give ratings of 9 or 10 to the best patterns you hear in the course of this experiment, and ratings of 1 or 2 to the worst, regardless of how you might judge these patterns in an absolute sense. Avoid giving too many ratings in the middle range; try to use the extremes as well...."

Nearly all subjects in fact used the whole range of rating categories.

Results and Discussion

Consistency of judgments. The first question to ask was whether the subjects were able to perform the task—that is, give reliable judgments. The reliability of their ratings could be determined by correlating the ratings across the three blocks of stimuli. Since the first block served to familiarize subjects with the stimuli, the correlation between the second and third blocks was expected to be higher than that between the first block and either of the other two. However, although this was true for some individual subjects, there was no such overall tendency in the data, and the three interblock correlations were therefore averaged for each subject.

All 14 musically experienced subjects exhibited significant average correlations, ranging from 0.31 ($p < .05$) to 0.79 ($p < .0001$). Of the 12 musically inexperienced subjects, however, only two showed a significant average correlation (0.49 and 0.51, respectively, both $p < .001$); for the rest, the correlations ranged from -0.01 to 0.18. This is a very striking difference. Most musically untrained subjects apparently did not possess a stable criterion by which to judge the stimuli.

A second criterion that separated the two subjects groups was their response to timing pattern Q20L. As was evident to the author during stimulus generation, this pattern (as well as Q40L) sounded really ridiculous, in contrast to the other patterns, which seemed at least moderately acceptable. Indeed, all musically

experienced subjects assigned their lowest ratings to Q20L, with average ratings ranging from 1.0 to 2.0. Ten of the 12 musically inexperienced subjects, however, gave this stimulus average ratings between 5.0 and 9.33! The remaining two subjects gave average ratings of 2.67 and 3.0, respectively; however, they were not the two individuals who showed significant reliability of judgments.

Because of this striking dichotomy in judgmental criteria and consistency between the two subject groups, further analysis was restricted to the data of the 14 musically experienced subjects. Their responses to the parabolic and hybrid patterns were analyzed separately.

Parabolic patterns. The parabolic patterns constituted a 3 (Type) by 5 (Curvature) design. The subjects' ratings were averaged over the three blocks and subjected to a two-way repeated-measures ANOVA. The average ratings are plotted in Figure 5.

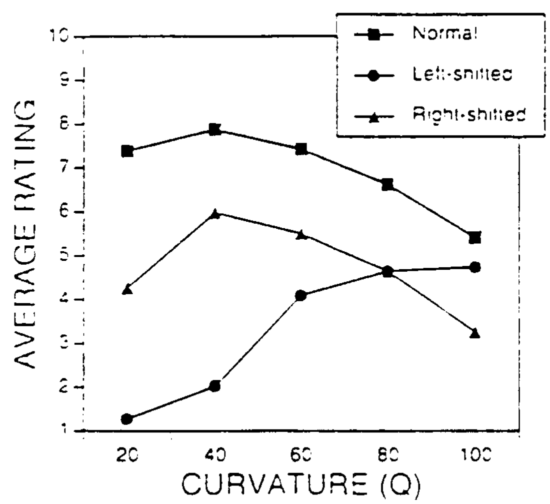

Figure 5. **Average ratings given to the parabolic patterns.**

As is evident from the figure, the prediction that the normal parabolic curves would receive the highest ratings was confirmed. The main effect of Type was highly significant [$F(2,26) = 65.60$, $p < 0.0001$]. There was also a significant main effect of Curvature [$F(4,52) = 5.50$, $p < 0.001$], though it was irrelevant in view of a strong two-way interaction [$F(8,104) = 20.95$, $p < 0.0001$]. This interaction was evidently due to the very different effect of Curvature for left-shifted parabolas than for normal and right-shifted ones.

The latter two stimulus types were analyzed in a separate ANOVA. There were significant effects of Type [$F(1,13) = 40.24$, $p < 0.0001$] and of Curvature [$F(4,52) = 11.26$, $p < 0.0001$], but a nonsignificant interaction [$F(4,52) = 1.44$, $p = 0.24$]. Normal parabolas were rated more highly than right-shifted ones at all degrees of curvature, and for both types the most preferred curvature was Q40. By contrast, left-shifted parabolas were judged extremely unfavorably at low curvatures (as noted earlier) and more favorably at high curvatures. At Q100, left-shifted functions were almost as acceptable as normal ones, and more acceptable than right-shifted ones. This indicates that the subjects were particularly averse to hearing a short first IOI; for stimuli Q60L to Q100L, the reduction of the starting tempo apparently compensated for the exaggeration of the final slow-down.

Hybrid patterns. The 30 hybrid patterns, together with the parent patterns Q20 and Q100, formed a $2 \times 2 \times 2 \times 2 \times 2$ design: Each of the five IOIs could either have a short duration (from Q20) or a long duration (from Q100). These five positions will be referred to by the letters A, B, C, D, E in the following. A 5-way repeated-measures ANOVA was conducted on the subjects' ratings. Significant main effects in this analysis would indicate that the listeners preferred a shorter or longer IOI duration in particular positions. Such effects were more likely in the positions where Q20 and Q100 differed most, i.e., E and A (cf. Figure 4). Of greater interest were any interactions among the five position factors, which would indicate that the relationships among several IOIs mattered. The average ratings are shown in Table 1.

Only one of the five main effects reached significance, that of position A [$F(1,13) = 8.10$, $p < 0.02$]: Listeners preferred the shorter IOI in that position (see Table 1, bottom row).[4] The main effect for the last position, E, was nonsignificant [$F(1,13)$ 1.06, $p < 0.32$], even though the change in duration was larger (424 ms vs. 264 ms). This is interesting in view of the "Cortot pattern" mentioned earlier, in which the last IOI is abnormally shortened; apparently, the present listeners were not very consistent in their responses to different degrees of final lengthening.[5]

Table 1. *Average ratings for the hybrid stimuli.*

Code	Rating	Code	Rating	Code	Rating	Code	Rating
H00000	7.4	H01000	6.5	H10000	5.4	H11000	6.9
H00001	7.6	H01001	6.5	H10001	4.9	H11001	5.1
H00010	6.8	H01010	6.1	H10010	5.2	H11010	5.6
H00011	6.0	H01011	6.1	H10011	4.6	H11011	5.3
H00100	7.1	H01100	6.7	H10100	5.7	H11100	5.5
H00101	7.0	H01101	6.1	H10101	5.5	H11101	6.0
H00110	6.8	H01110	6.6	H10110	4.7	H11110	5.7
H00111	6.8	H01111	6.8	H10111	4.2	H11111	5.4
H00 ...	6.9	H01 ...	6.4	H10 ...	5.0	H11 ...	5.7
H0 ...	6.7	H1 ...	5.4				

There were several significant interactions, however, which indicated that listeners did not judge IOI durations individually. The three largest interactions were AB [$F(1,13) = 16.59$, $p < 0.002$], ABCD [$F(1,13) = 16.15$, $p < 0.002$], and CE [$F(1,13) = 12.00$, $p < 0.005$]. Four additional interactions, ACD, BCD, BD, and BDE, were significant at $p < 0.05$, and three further interactions, ACDE, ABCE, and BCDE, were nearly significant ($p < 0.06$). It is perhaps noteworthy that the only two positions that were never involved together in a significant interaction are A and E. The beginning and the end of the timing pattern thus seemed to be judged independently.

These interactions indicate that it is the pattern of IOIs that mattered, not individual IOI durations. The AB interaction, for example, shows that a shorter second IOI (B) was preferred when the first IOI (A) was short, but a longer B was preferred when A was long (see Table 1, penultimate row). The BD and CE interactions show a similar pattern of preferred positive covariation between two positions. Now consider a more complex interaction, ABCD, which subsumes two other significant interactions, ACD and BCD. It can be viewed as four CD interactions, one for each of the four combinations of A and B values. Three of these four two-way interactions exhibit the positive covariation described above, but one (that for long A and short B) exhibits negative covariation. That is, in that specific condition listeners preferred a long C when D was short, and a short C when D was long. The reason for this complex interaction is not obvious, but it is remarkable that position C, which had a duration difference of only 24 ms, was so strongly involved in it. Listeners' sensitivity to small deviations in that position mirrors the restricted range of observed IOI durations.

Another prediction may be examined in the hybrid pattern data. The parent patterns, Q20 and Q100, were parabolas of the normal type. The hybrid patterns approached parabolas in various degrees. Therefore, none of them should have been rated higher than the parent patterns, whereas quite a few should have been rated lower. However, the difference in average ratings between Q20 and Q100 of about 2 points (cf. Figure 5) must be taken into account. The revised prediction, therefore, is that no hybrid pattern should have been judged more acceptable than Q20, but some should have been judged less acceptable than Q100.

The first part of the prediction was confirmed: Only one hybrid stimulus, H00001 (i.e., Q20 with a lengthened final IOI), received a higher average rating than Q20 (7.6 vs. 7.4), but this difference was certainly not significant. The second part of the prediction was also supported: There were 7 hybrid stimuli that received lower average ratings than Q100 (5.4). The lowest rated stimulus, H10111 (4.2), corresponded to Q100 with a shortened second IOI. In fact, Table 1 shows that all stimuli of the type H10... received low ratings whose range (4.2 to 5.7) did not overlap at all with the ratings of H00 . . . and H01 . . . stimuli (range: 6.1 to 7.6); H11 . . . stimuli were in between (range: 5.1 to 6.9). This again confirms the relative importance of the first IOI in relation to the second. Clearly, listeners did not like a relatively long first IOI. This is easily explained by the tonal structure: The first tone of the melodic gesture, E, is a half-step below the tonic (i.e., a "dissonant lower neighbor") and, moreover, in a metrically weak position, which calls for a quick resolution.

It might also be asked whether the ratings of the hybrid stimuli reflected the degree to which they approached a parabolic timing curve. As the results for parabolic patterns show, however, what matters is not so much the parabolic shape itself as its parameters. That deviations from a parabolic shape can be tolerated is illustrated by the ratings for hybrid stimulus H11110 (5.7), which were slightly higher than those for the perfectly parabolic stimulus H11111, alias Q100 (5.4). H11110 resembles the "Cortot pattern," and Cortot may have taken advantage of listeners' tolerance for variations in the final IOI. The resulting timing shape is only moderately acceptable, however, which matches the author's impression from listening to Cortot's recordings (whatever other qualities they may have).

GENERAL DISCUSSION

The present results are limited in a number of ways, which will be discussed below. Within these limitations, however, they provide a clear indication of a constraint on performance timing that is shared by expert performers and musically acculturated listeners. While it may be perfectly obvious to some theorists that such constraints must exist, their objective demonstration and characterization has rarely been undertaken before. The local constraint examined here is flexible enough to permit a large variety of concrete timing patterns; yet there is reason to believe that, in a specific musical context, a single pattern may be optimal. Because of the contextual timing variation inherent in different performances, the evidence for optimality comes from the perceptual data alone. For the specific musical excerpt presented here, the timing shape labeled Q40 seemed to be best, on the average.

It is necessary to discuss now what possible generality this finding may have. Three major issues concern individual differences in preferences and experience, the specific stimulus conditions of the experiment, and the specific musical excerpt selected.

Individual differences among listeners did exist, of course, as they do in nearly all psychological studies. However, the high levels of significance of some of the effects obtained suggest considerable agreement. More extensive replications of judgments per subject would be needed to interpret individual differences. A few observations are offered here: All subjects but one gave some of their highest ratings to parabolic patterns of the normal type; the exception was a professional pianist who gave her highest ratings to stimuli H11110 (the Cortot-like pattern), H11010, and Q40R, which all shared an initial *accelerando* but had a reduced *ritardando* at the end. Among the normal parabolic patterns, most subjects' preference fell on patterns with lower curvature (Q20, Q40, or Q60), though two subjects, both accomplished pianists, preferred those with higher curvature. One subject, interestingly the youngest in the group (an 11-year old girl who studies the piano), did not differentiate much among the different degrees of curvature, though she clearly preferred the normal patterns over the left- and right-shifted ones. How the internal standards by which such patterns are judged are acquired in the course of music education is of course a very interesting question for future research.

That musical experience is a *sine qua non* for reliable performance in the experimental task was demonstrated convincingly here. The precise nature of the necessary experience is less clear, however. The subject sample did not include individuals who cannot play an instrument but listen extensively to classical music; the musically experienced subjects were all instrumentalists of varying degrees of proficiency. The several professional pianists in the group, who surely had the most extensive musical education, actually were not the most reliable judges. It is entirely possible that professional musicians' criteria are less fixed than those of amateurs and ordinary music lovers, because constant interaction with other musicians as students, ensemble players, and teachers may encourage tolerance of a large variety of interpretative nuances.

It seems unlikely that specific knowledge of "Träumerei" and exposure to performances of this music in the past played a significant role in subjects' judgments. Most subjects were very familiar with the piece, but some were not. One subject, a flutist, indicated that she did not know it at all; two others, who are string players, and the 11-year-old pianist indicated they were "fairly" familiar with it. Yet, these subjects gave reliable judgments consistent with those of the other musicians. The more important argument is one of plausibility: Although some surface characteristics of previously heard performances may well be part of the memories of familiar pieces of music, performance rules must also be stored in a more abstract form, so as to be applicable to music never heard before. Musically experienced listeners surely can judge the performance quality of novel music in a familiar style, just as performers can sightread new music with good expression, again

provided the style is familiar (as it is in the case of any piece from the Romantic period).

Subject variables thus do not seem to impose a serious limitation on the generality of the present results. Stimulus variables are more of a problem. There are at least three factors that may influence subjects' judgments but were kept constant in the experiment. One is the melodic and harmonic content of the musical excerpt. As pointed out in the section on performance measurements, the melodic gesture under examination occurs six times in "Träumerei," and only two of these instances are exactly identical. As Figure 2 showed, the average timing curve of the excerpt used in the rating task (which occurs in bars 1-2 and 17-18) has only moderate curvature, comparable to the Q40 stimulus. A lower curvature was typical of the variants in the middle section of the piece, while high curvatures were mainly associated with the last instance preceding the *fermata*. Thus the listeners indeed preferred the curvature appropriate to the excerpt offered, but they may well prefer a different curvature for other variants. However, the preference for normal parabolic shapes should hold across all variants.

A second factor is the timing (and the implied tempo) of the context in which the critical melodic gesture was presented. The IOIs of the preceding musical events were set somewhat arbitrarily at the geometric means of the performance sample. It is possible, even likely, that a different choice of IOIs for the context would have influenced subjects' preferences. For example, if the IOIs had been longer (implying a slower tempo), listeners may have opted for a more curved or elevated timing function. This would be interesting to test in future experiments. As it was, however, listeners were presented with an average contextual timing pattern, and they preferred a curvature that also corresponded to the average, which seems appropriate. Their general preference for normal parabolas should be independent of variation in contextual timing.

A third factor is the intensity microstructure of the melodic gesture, which was also held fixed. It was derived from an individual performance, and its contour may not have been close to the average.[6] It did constitute a *crescendo*, however, as marked in the score. Very little is known at this time about the perceptual interdependence, if any, of timing and intensity microstructure. It is conceivable that a different intensity contour would change subjects' curvature preferences. Again, however, there is no reason to believe that the subjects would prefer atypical timing patterns, as long as the intensity microstructure stayed within the normal range of variation.

A final consideration is the selection of timing functions presented in the experiment. Clearly, there are many possible shapes that were not included, mainly because they were expected to sound terrible and might have offended musical listeners' sensibilities. This is not a serious omission. On the other hand, it is conceivable that there are timing curves superior to Q40 in this particular context. The left- and right-shifted parabolas constituted fairly gross deviations, and there are other functions closer to the normal ones that, in a sufficiently sensitive perceptual test, might prove even more highly acceptable. It must also be noted that implicit tempo (which is difficult to quantify in a temporally modulated performance; see Repp, 1992) was confounded with curvature to some extent, Q100 having a slower tempo than Q20. Listeners' overall preference for Q40 may have constituted a preference for (contextually appropriate) tempo as much as for curvature. This would have to be sorted out by varying the constant and quadratic parameters of the timing curves independently.

In summary, consideration of various stimulus-related factors suggests that listeners' preference for a particular curvature of the timing function may well be context-dependent; however, their general preference for normal parabolic shapes most likely is not. It should also be remembered that the normal family of parabolic shapes was derived from a set of performances that varied widely in the performance parameters (tempo, contextual timing, intensity microstructure) whose possible role in perception was just considered. The generality of the parabolic constraint across this performance variation should have a parallel in perceptual preference across similar variation.

This leads to the broader question concerning the generality of the parabolic constraint to other kinds of melodic gestures and musical styles. One obvious limitation is that the constraint can meaningfully apply only to melodic gestures that have at least four IOIs. The more IOIs, the stronger the constraint may manifest itself. Repp (1992) examined the timing patterns of three other melodic gestures in "Träumerei," each comprising 4 IOIs near the end of a phrase; they, too, seemed to follow the constraint, but somewhat less consistently than the 5-IOI gesture examined here. Gestures with less that 4 IOIs, of course, cannot violate the parabolic constraint; they are simply irrelevant to it.

Another limitation is that the gestures may need to have a *ritardando* in them. This was true for all the instances examined by Repp (1992). Moreover, the present results are in strong agreement with the performance and perception results of Sundberg and Verrillo (1980), who focused on final *ritardandi* in Baroque music. The parabolic constraint thus may characterize *ritardandi* at all levels of the grouping structure, and quite possibly across different musical styles. It may indeed represent a "natural" way of changing tempo, including both *accelerando* and *ritardando*, though the evidence for *accelerando* is limited to the initial part of the melodic gesture examined here.

These tempo changes, moreover, must be uninterrupted. This perhaps constitutes the most serious limitation of the parabolic constraint. It may only apply to gestures that are rhythmically uniform and do not contain tones that receive special emphasis for harmonic or melodic reasons. If so, it characterizes only a small minority of the melodic gestures in a musical piece, though they may be the most salient ones, which mark the ends of major sections. This minority, however, turns into a majority if all short melodic gestures in which the constraint applies trivially are included. It is noteworthy that Todd (1992), in the process of extending his coarse-grained model of expressive timing (Todd, 1985) to detailed local timing patterns, has been assuming a linear velocity function of tempo change for melodic gestures ("segments") of any length, apparently with good success. A linear velocity change is equivalent to a quadratic timing function for the raw IOIs during a unidirectional tempo change. Previously, Todd (1985, 1989) presented data suggesting that the global timing shapes of whole phrases can be modelled by a family of parabolic functions. His current, somewhat modified conception promises to contitute a valid basis of a general performance model.

The parabolic functions used in Repp (1992) and in the present study were empirically derived and may eventually have to give way to similar but theoretically motivated functions such as proposed by Todd (1992), provided that they fit the data equally well. The extramusical origin of constraints on performance timing is still a matter of speculation, but it is likely to lie in aspects of physical movement that have invaded musical performance and ultimately account for the frequent allusions to "musical motion" in the musicological literature. Although musical motion is often attributed to tonal sequences without explicitly appealing to performance, it seems likely that music needs to be *set into motion* by a performer, real or imaginary. Once the physical movement has entered the music, it will in turn be able to "move" a listener, provided it has the properties that the sensitive listener is attuned to. The kinds of melodic gestures that are most "moving" in a good performance are probably those that give the timing constraint a chance to emerge clearly and impress itself on the listener.

REFERENCES

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal*, 15, 56-71.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology*, 24, 193-213.

Kronman, U., & Sundberg, J. (1987). Is the musical ritard an allusion to physical motion? In A. Gabrielsson (Ed.), *Action and perception in Rhythm and Music* (pp. 57-68). Stockholm: Publications issued by the Royal Swedish Academy of Music No. 55.

Lindblom, B. (1978). Final lengthening in speech and music. In E. Garding, G. Bruce, and R. Bannert (Eds.), *Nordic Prosody* (pp. 85-101). Lund University, Sweden: Department of Linguistics.

Lussy, M. (1882). *Musical expression: accents, nuances, and tempo in vocal and instrumental music*. London: Novello.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331-346.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, 88, 622-641.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America*, 92(5), 2546-2568

Seashore, C. E. (1938/67). *Psychology of music*. New York: McGraw-Hill/Dover.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, 13, 326-376.

Sundberg, J., Friberg, A., & Frydén, L. (1991). Threshold and preference quantities of rules for music performance. *Music Perception*, 9, 71-91.

Sundberg, J., & Verrillo, V. (1980). On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America*, 68, 772-779.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-58.

Todd, N. (1989). A computational model of rubato. *Contemporary Music Review*, 3, 69-88.

Todd, N. (in press). The kinematics of musical expression. *Journal of the Acoustical Society of America*.

FOOTNOTES

**Music Perception*, in press.

[1] These weaknesses include preselection of performances with "typical" *ritardandi*, averaging across heterogeneous musical materials, a rather unbalanced and poorly described design in the perception experiment, and great variability in listeners' judgments.

[2]That is, as long as the pattern does not lead musically literate listeners to conclude that a rhythmic *mistake* has been made. In other words, the performance timing pattern must be compatible with the notated temporal pattern. There is, of course, a grey area here in which faithfulness to the score may be a matter of opinion.

[3]The remaining two exceptions, both occurring in the performance of Brazilian pianist Cristina Ortiz, exhibited a relative lengthening of the third IOI instead (i.e., a W-shaped pattern).

[4]This seems to contradict the earlier observation that subjects disliked a short first IOI in left-shifted patterns. Note, however, that the first IOI of stimulus Q20 corresponded to that of stimulus Q80L (cf. Figure 4). Thus, while listeners disliked abnormally short first IOIs, within the normal range they preferred short over long first IOIs.

[5]It should not be inferred that the subjects were unable to detect the difference in duration of the final IOI (424 ms). The present task was one of aesthetic, not sensory discrimination.

[6]The methods of deriving and transferring the intensity values will not be defended here, as they are still in need of validation. Suffice it to say that the dynamic variation sounded appropriate to the author. An analysis of the intensity microstructure of the entire sample of 28 "Träumerei" performances remains to be conducted.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 277-282

A Review of Carol L. Krumhansl's *Cognitive Foundations of Musical Pitch**

Bruno H. Repp

The psychology of music perception and cognition, nearly dormant 15 years ago, has made considerable strides in the last decade. Several textbooks and edited collections of articles have appeared, two new journals have been established, societies have been formed, and many reports of empirical studies have been published, including one in monograph form (Serafine, 1988). However, the field is still new and small,[1] and few researchers have had the persistence and the good fortune of continuous grant support to develop and bring to fruition an extended and coherent program of research. Carol Krumhansl, of Cornell University, has accomplished that feat, and her monograph documents a decade of individual achievement, resulting in a critical mass of psychological data organized in a tight conceptual framework. This publication is a landmark event in a young field striving for definition and recognition.

The brilliance of Krumhansl's approach was recognized early on by her peers who bestowed on her the American Psychological Association's Distinguished Scientific Award for an Early Career Contribution in Psychology (see American Psychologist, March 1984, pp. 284-286). The citation honored her for "a dazzling interplay of experimental techniques, music theory, and multidimensional scaling" that had uncovered "new cognitive structures of great richness and beauty" (ibid., p. 284). This methodological virtuosity as well as its satisfying results are evident throughout the book. Although nearly all the results have been published previously in accessible journals, they are brought together here for the benefit of the reader, who is led through the complex issues by lucid explanations and discussions. The clear organization and sense of direction make reading the book an aesthetic as well as an intellectual pleasure.

The book is divided into 11 chapters. Chapter 1 introduces the reader to the author's objectives and methods. Some general remarks about the approach of cognitive psychology are provided for nonpsychologists. The general aim is "to describe the human capacity for internalizing the structured sound materials of music by characterizing the nature of internal processes and representations" (p. 6). The more specific aim of Krumhansl's research is "to describe what the listener knows about pitch relationships [mainly in traditional Western music], how this knowledge affects the processing of sounded sequences, and how this system arises from stylistic regularities identifiable in the music" (p. 9). Krumhansl's distinctive way of characterizing internal representations is to depict the similarity relationships within sets of basic elements as distances in a multidimensional space.

The basic elements are said to be single tones, chords, and keys. (That keys are rather more abstract entities than are tones and chords is not immediately pointed out, but perhaps obvious.) Krumhansl does not further defend this axiom, which would be accepted by most music psychologists and musicologists. Witness, however, what Serafine (1988)—not cited in the book—had to say: "On this view, the elements and processes of cognition will be exactly isomorphic to the factors we are able to find ... and manipulate in experiments" (p. 26) and "we know that the stimuli used in such studies are never, under any circumstances, considered or listened to as music" (p. 25). And, further along, Serafine argued that "much psychological research has mistakenly focused exclusively [on], and also misinterpreted, merely the results of reflection—that is, scales, chords, and discrete pitches—rather than been concerned with music itself" (p. 52). I will return to these arguments at the end of this review.

Chapter 2 introduces the reader to the concept of tonal hierarchy, and to Krumhansl's way of deriving and depicting this cognitive structure. The term "hierarchy" here refers to a simple ordering of tones according to their relative importance or stability within a given key, not to a structure with several nested levels (as in Lerdahl and Jackendoff, 1983). The hierarchy of the tones within a given key is likened to the organization of category members around a prototypical exemplar, in this case the base note or tonic, which serves as a cognitive point of reference. Krumhansl's experimental probe tone method (developed in collaboration with Roger Shepard) presents listeners with a sequence of notes that unambiguously define a particular key (e.g., a scale or a tonic triad chord), followed by a single note of variable pitch. Subjects judge on a rating scale how well this final probe tone fits into the context of the established key. The resulting pattern of average ratings across all tones of the chromatic scale describes the tonal hierarchy: The tonic (scale step 1) is rated highest, followed in major keys by scale step 5, steps 3 and 4, steps 2 and 7, and finally the chromatic tones that are not members of the key. In minor keys, the order is 1, 3, 5, then steps 2, 4, and 7, and finally the chromatic tones. These hierarchies correspond to the functional importance of the scale notes in traditional tonal music, as described by musicologists.

By computing the auto- and cross-correlations between the rating profiles for all possible pairs of major and minor keys, Krumhansl derives a matrix of interkey similarities that she then subjects to nonmetric multidimensional scaling to obtain a spatial configuration of interkey distances. The configuration is strikingly regular, due to the constraints built into the data, and it also makes sense: Two dimensions in which the points representing the keys are arranged according to the circle of fifths are convolved with two dimensions in which the keys are arranged in a circle that reflects relative and parallel relationships between major and minor keys. The total four-dimensional pattern can be visualized as the surface of a torus (a doughnut), or the surface can be spread out in two dimensions representing the angular coordinates of the keys in the two circular configurations. This latter, two-dimensional key map resembles maps drawn intuitively by musicologists: The key of C major, for example, is adjacent to the major keys differing in one note (G and F major), to the relative minor key (a minor), and to the parallel minor key (c minor). Thus it provides an empirical validation of musicologists' insights through listeners' probe tone ratings.[2] Krumhansl adds a cautionary note by pointing out that her model does not account for possible directional asymmetries in key similarity.

Following this methodological tour de force, the author turns in Chapter 3 to a discussion of the factors that may underlie listeners' knowledge of tonal hierarchies. She considers two: the phenomenon of tonal consonance, and the statistical distribution of pitches in tonal music. Strong correlations of tonal hierarchies with consonance hierarchies would suggest that tonal hierarchies originate in the acoustics of complex tones and therefore are relatively fixed and universal. Stronger correlations with the distribution of tones in familiar music, on the other hand, would suggest that tonal hierarchies are learned and culture-bound. Krumhansl briefly reviews acoustically-based theories of tonal consonance and then proceeds to describe the correlations between her tonal hierarchies and consonance hierarchies culled from various studies in the literature.[3] The correlations are moderately high for major keys but lower and mostly nonsignificant for minor keys. Krumhansl then proceeds to compare the tonal hierarchies with the statistical frequency distributions of tones in various selections of tonal music, again obtained from the literature. These correlations are much higher and significant for both major and minor keys. Finally, a multiple regression analysis is performed which demonstrates that tonal consonance does not account for any aspect of tonal hierarchies that is not also accounted for by tonal frequency distributions. On the basis of these results, Krumhansl argues that tonal hierarchies are learned through listening to tonal music and hence are a product of musical acculturation. Research by Lynch et al. (1990a, 1990b)—too recent to be cited by Krumhansl—indeed suggests that this acculturation begins in the first year of life.

In Chapter 4, Krumhansl turns to a practical application of her tonal hierarchy results: determination of the key for a musical excerpt, and of changes in key as music progresses. Her key-finding algorithm (developed with Mark Schmuckler) is simple: The total duration of each note in the musical excerpt is determined by combining repeated occurrences of the same note, regardless of octave or ordinal position. The resulting relative durations of the 12 notes in the octave (with zero for notes that do not occur) are then correlated with the tonal hierarchy profiles for the 24 major and minor keys, as described in

Chapter 2. The largest correlation identifies the dominant key. Other large correlations identify related keys that may also be suggested by the musical passage. Indeed, the major virtue of the algorithm is seen in its ability to yield a key hierarchy, rather than just a single dominant key. As Krumhansl demonstrates, music theory experts can rate the relative strengths of candidate keys for short musical excerpts.

The effectiveness of the key-finding algorithm is demonstrated in three specific applications and is compared to other procedures proposed in the literature. In the first application, the algorithm is used to determine the nominal keys of preludes (24 each) by Bach, Shostakovich, and Chopin, based on the first four notes only. The results are quite accurate for Bach and Shostakovich, but less so for Chopin. In the second application, 24 fugue subjects of Bach and Shostakovich are analyzed in terms of how many notes are needed to determine the correct key. The average number of notes is about 5, considerably less than required by alternative algorithms proposed in the literature. In the third and most elegant application, the key modulations in a single Bach prelude are tracked measure by measure and compared to judgments by two experts. There is good agreement, though the algorithm does not quite match the experts. The key changes are represented graphically as a path on the surface of the torus representing the configuration of interkey distances (Chapter 2). A final section of Chapter 4 acknowledges the current limitations of the algorithm, which include its insensitivity to temporal order, melodic patterns, harmonic structure, and rhythmic stress.

In Chapter 5, Krumhansl returns to perceptual data and in fact reports an original study not published elsewhere, which replicates and extends one of her early experiments. The topic is the perceived relation between two musical tones. Whereas in the experiments that led to the tonal hierarchy profiles the subjects' task was to judge how well a single note fit into the tonal context established by a precursor sequence, listeners now hear two notes following the key-defining context, and the task is to rate how well the second note goes with the first. The goal is to demonstrate that these perceived relations depend on the tonal context—for example, that the notes C-G are perceived as a better sequence than the notes C#-G# when the key is C major, even though both note sequences represent the same musical interval (a fifth). Krumhansl starts out by discussing various spatial representations of the psychological pitch relations among tones, which increasingly take the functional roles of tones into account. Although the author persists in talking about the perceived relations among successive tones, what her experiment is really about is the functional role of two-note sequences within a given key. It comes as no surprise, then, that the order of the two notes plays an important role, a factor that cannot be accommodated by traditional multidimensional scaling of similarity data. Krumhansl nevertheless presents the results of such an analysis, but also notes its shortcomings. The spatial solution that best approximates the perceived tonal relations shows the tonic at the vertex of a cone, along whose circumference the other tones are arranged according to pitch, but with their distance from the tonic being an inverse function of their position in the tonal hierarchy. A more complete picture including order effects emerges from a multiple-regression analysis: Listeners' judgments were most strongly influenced by the position in the tonal hierarchy of the second tone, with weaker but significant contributions of the tonal hierarchy of the first tone, the pitch distance between the two notes, and the distance between the two notes along the circle of fifths.[4] The chapter concludes with a demonstration that the results are positively correlated with the relative frequencies of melodic intervals in several musical corpora, as tabulated previously by others.

Chapter 6 first summarizes three principles that have emerged from this research and from the work of others on perceptual organization and memory. The principle of contextual identity states that stable tones (i.e., tones high in the tonal hierarchy) are remembered better than unstable tones. The principle of contextual distance states that two tones are perceived as the more closely related (and hence are also more easily confused in memory) the more stable either of them is. The principle of contextual asymmetry states that two tones are perceived as more closely related when the second tone is more stable than the first than when they are in the opposite order. These principles are expressed formally in terms of perceptual distances, and relevant findings are cited from the literature. The principles are said to support basic tenets of Gestalt theory, with tonality providing a kind of Gestalt quality, though (to this reader) this argument does not add any explanatory power. The second half of the chapter discusses perceptual grouping principles in music, with data from several recent studies by the author and her collaborators. These studies show that pitch and rhythm make independent

contributions to perceived phrase structure, that there are reliable boundary cues in performances of pieces by Mozart as well as Stockhausen, and, most intriguingly, that 6-month old infants prefer music that is interrupted at phrase boundaries to music that is interrupted in the middle of a phrase. Lowering of pitch and increases in tonal duration are identified as boundary cues likely to have been salient to these infants, and the analogy to speech prosody is noted.

Chapters 7 and 8 are easily summarized. They report the results of experiments with chords that replicate in all essentials the experiments with tones described in Chapters 2 and 3. Chapter 7 reports data not published previously. Listeners were presented with one (Chapter 7) or two (Chapter 8) triadic chords following a key-establishing context and judged how well they followed the preceding context. The results are shown to reflect the relative stability of the chords in the tonal system, and they illustrate each of the three general principles discussed in Chapter 6. Memory for chords in a sequence is also shown to reflect relative stability, and chord stability is found to correlate with frequency of occurrence of chords in tonal music. Krumhansl concludes by summarizing the many parallels between the perceptual organization of tones and chords.

All the work up to this point can be considered as concerned with establishing basic facts concerning tonal organization in perception and memory. In Chapter 9, Krumhansl summarizes two studies that make use of these basic data in addressing two more complex scenarios: key modulation and polytonality. In the key modulation experiment, probe chords are presented after every single chord of chord sequences that modulate to close or distant keys. By correlating subjects' judgments with the tonal hierarchy profiles obtained previously in unambiguous key contexts (Chapter 2), the relative strengths of different keys can be assessed as the chord sequence unfolds. By treating these strengths as distances, the changing sense of key through the chord sequence can be represented as a path in the toroidal key-distance map derived in earlier studies. The analysis reveals listeners' initial resistance to radical key changes, followed by abrupt shifts into the new key when the following context confirms it. In the experiment on the perception of polytonality, a famous excerpt from Stravinsky's "Petrouchka" is used in which two distant keys (C# and F# major) are used simultaneously.[5] Probe tones are presented after the bitonal passage, as well as after each tonal component played separately. Detailed analysis of the results suggests that subjects' judgments are governed by the notes presented, and hence also by both keys, but that no clear sense of either tonality develops. Listeners were generally unable to focus on one or the other tonality, even when instructed to do so. Thus it seems that polytonality, in this instance at least, prevents the establishment of either a single or a multiple tonal framework; instead, it creates ambiguity.

The author ventures farther afield in Chapter 10, which reports studies that applied the probe tone technique to 12-tone serial music, to North Indian classical music, and to Balinese gamelan music (the last study done by Kessler and colleagues). The resulting probe tone profiles, obtained at various points during and/or following musical excerpts from these various styles, were analyzed to determine the factors that played a role in subjects' judgments. In the study of 12-tone music (excerpts from two of Schoenberg's works), two groups of subjects could be distinguished whose patterns of responding were almost exact opposites of each other: One group, generally more familiar with 12-tone music, avoided tonal implications like the plague, whereas the other group was governed by whatever tonal implications they could derive from surface features such as note length and recency. Similarly, in the experiment using North Indian music in different keys (thats), experienced subjects gave probe tone profiles that enabled Krumhansl to recover through multidimensional scaling analysis the key (that) distance map postulated a priori, whereas other subjects gave a much less clear pattern and seemed to be governed by surface features of the music rather than by the underlying scales. Krumhansl's conclusion that "listeners can set aside ... expectations and hear the pitch events in style-appropriate terms quite independently of their prior musical experience" (p. 268) is perhaps premature, but her results demonstrate that musical compositions in different styles often provide the "surface" information (emphasis, repetition, lengthening of important notes) a listener needs to infer the characteristics of the style, so the prior experience is simply not needed to appreciate simple structural features. Interestingly, orthodox 12-tone music is different in that it studiously avoids such surface aids to the listener, so, in order to respond appropriately to this music, listeners need to know what not to expect. This is an interesting demonstration of the inherent radicalism of dodecaphonic music, and an indication that it negates not only traditional (i.e., 19th century) aesthetic values but psychological principles as well.

In her final chapter, Krumhansl first discusses rather briefly some formal properties of the tonal system and of some other scale systems,[6] and speculates that these properties may have arisen from psychological constraints, thereby suggesting interesting future research to be done. The final pages summarize the principal findings from the empirical studies. The perception of tonal music is said to exhibit "one of the hallmarks of a cognitive system: the categorization and classification of sensory information in terms of a stable, internal system of structural relations" (p. 282). That system, Krumhansl claims, is abstracted and internalized by listeners from the sound events in the music they encounter; that is, it is learned and style-specific, though it makes use of general cognitive architecture to represent the external regularities.

Krumhansl's book is a superb accomplishment and represents cognitive psychology at its best. This does not mean that it is beyond criticism. The question is, quite simply, whether cognitive psychology at its (current) best is good enough to explain musical phenomena. Music is a very highly developed art form whose complexities have kept musicologists busy for centuries. Cognitive psychology is not particularly well suited to studying art forms, or at least has not yet proven to be. What, to Krumhansl, are major insights gained from a decade of research may be platitudes to a musicologist (Butler, 1990) or musician. This problem is endemic to cognitive psychology, which searches for general principles that cut across many domains. However, it is with the specific properties of music that the study of music proper begins. It could be argued that cognitive psychology and the serious study of music are mutually exclusive, though perhaps complementary. If so, then even a tour de force such as Krumhansl's research will inevitably miss the significant issues in music perception. Nevertheless, it may provide a general framework within which these music-specific issues may be addressed in a more rigorous manner.

The probe-tone task has been criticized by Butler (1989) as being insensitive to the dynamic unfolding of harmonic implications in tonal music, as permitting alternative listener strategies, and as being more sensitive to the tone distributions in the key-defining context than to the implied tonality. Krumhansl's reliance on tabulations of note durations and frequencies was likewise attacked by Butler as being a crude method. The resulting exchange (Krumhansl, 1990; Butler, 1990) has not settled these issues completely, and further research will be necessary. It certainly would be inappropriate to conclude (as Butler tends to do) that any of Krumhansl's results are artifactual until they have been proven to be so by careful follow-up experiments.[7] For one thing, most of her findings are in good agreement with conventional musical wisdom, which makes it likely that they will stand the test of time. It seems to this reviewer that Krumhansl has justifiably ignored some significant musical detail in order to arrive at generalities, but the detail will have to be dealt with eventually. The dynamic tracing of harmonic expectations described in Chapter 9 certainly is an interesting beginning in that direction. Unfortunately, the probe tone method becomes prohibitively time-consuming as a tool for investigating modulations in real music, and trained musicologists' judgments may ultimately prove not only more convenient but also more reliable.

Like most cognitive psychology studies, Krumhansl's research is not concerned specifically with expert knowledge. After finding early on that musically untrained subjects tend to use nonmusical response strategies in the probe tone task, she relied in the following on listeners who had considerable musical training but were not necessarily professional musicians or musicologists. This is both a strength and a weakness. It is a strength in so far as it demonstrates the solid, ingrained knowledge musically informed listeners have of the tonal system. It is a weakness in that it does not characterize what, if anything, musically uninformed listeners know about music, and, more importantly, in that it misses the special skills and insights provided by highly trained musicians and musicologists. In any investigation of a highly developed art, expert judgments must be the measure of validity—even if those judgments sometimes diverge. The consensus of average listeners can tell us what the average listener knows, but it will not capture the full subtlety of the phenomena under investigation.

What about Serafine's (1988) warnings, cited at the beginning of this review? It is certainly true that Krumhansl took certain musical elements—the "results of reflection"—as given and proceeded to develop her representations of mental structures in terms of those units (tones, chords, and keys). Her claim is undoubtedly that, even when not reflected upon, these units play a functional role in mental processing. It is also true, however, that the probe tone task directs the listener's attention to a particular unit (a tone or chord) and requests a judgment about it in the context of an

often stereotyped and much-repeated, musically trivial context that, moreover, is rendered in electronic sound and with mechanical timing and dynamics. There are exceptions, such as the experiment using actual excerpts from Schoenberg's music as the context for probe tones (Chapter 10). On the whole, however, it is quite possible that the musical samples in Krumhansl's experiments were not "listened to as music," by which Serafine (1988) presumably means that they were not perceived as musically meaningful or expressive. It is the more remarkable, then, that these meaningless sequences inevitably and strongly engaged the listeners' knowledge of tonal hierarchy structures; in a sense, then, they had some musical meaning, after all. It is the cognitive psychologist's trump card that even highly schematic stimuli often engage mental structures designed for much more complex and ecologically valid events. It is the expert's wild card, however, that only a very limited subset of pertinent structures can be probed in this way, so that an impoverished view of complex phenomena may result.

In her introduction, Krumhansl refers several times to "musical experience," but her research does not really deal with listeners' experiences. They made judgments that followed certain patterns; what they experienced, we do not know—probably boredom. Krumhansl's spatial maps present us with crystallized configurations—mental structures in vitro, as it were, that can be regarded with awe, like a piece of modern architecture. They convey none of the excitement and pleasure that comes from exploring the building, its corners and hallways. For an appreciation of musical meaning, we must read Langer (1953) or Zuckerkandl (1956) or Clynes (Clynes & Nettheim, 1982)—or simply listen. Krumhansl's cognitive world is one of discourse about music, not of music as "significant form" (Langer, 1953). Yet, there must be a close relation between the two. The characterization of that relation is perhaps the fundamental problem of music psychology. Krumhansl would be well equipped to tackle it as the next step in her remarkable career.

REFERENCES

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception, 6*, 219-242.

Butler, D. (1990). Response to Carol Krumhansl. *Music Perception, 7*, 325-338.

Clynes, M., & Nettheim, N. (1982). The living quality of music: Neurobiologic patterns of communicating feeling. In M. Clynes (Ed.), *Music, mind, and brain. The neuropsychology of music* (pp. 47-82). New York: Plenum Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* Oxford Psychology Series No. 17. New York: Oxford University Press.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception, 7*, 309-324.

Langer, S. (1953). *Feeling and form.* New York: Chales Scribner's Sons.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990a). Innateness, experience, and music perception. *Psychological Science, 1*, 272-276.

Lynch, M. P., Eilers, R. E., & Oller, D. K. (1990b). Musical acculturation in the first year of life. *Journal of the Acoustical Society of America, 87*, S19 (Abstract).

Serafine, M. L. (1988). *Music as cognition. The development of thought in sound.* New York: Columbia University Press.

Zuckerkandl, V. (1956). *Sound and symbol. Music and the external world.* Princeton, NJ: Princeton University Press.

FOOTNOTES

**American Journal of Psychology, 104*, 611-621 (1991).

[1] That is, considered as a post-war empirical enterprise. The psychology and philosophy of music have, of course, a distinguished history that goes back many centuries.

[2] Interestingly, the map goes beyond earlier representations in that it suggests that C major is closely related to yet another key, e minor—which, in its descending melodic version, differs in just one note from C major—but not to d minor, which also differs in one note. It is not clear whether this observation is substantiated by any musicological evidence.

[3] A consonance hierarchy—my term—results from quantitative estimates of predicted or perceived consonance for all tones of a scale when they are sounded together with the tonic of that scale.

[4] It is possible to regard the two-tone judgment task as a version of the one-tone task: The first probe tone merely extends or perturbs the tonal context in which the second tone is judged.

[5] An unfortunate mistake in this otherwise very carefully edited volume occurs in the musical examples on page 229: The bottom staves should be in treble clef throughout, not in bass clef. On the same page, in the penultimate line, "diminished triads" should read "diminished chords". Also, Krumhansl's spelling of "Petroushka" is an unfortunate amalgam of Stravinsky's original French "Pétrouchka" and its anglicized version, "Petrushka".

[6] An error occurs on page 277: There are not two but three octatonic scales; the "2-scale" was mistakenly omitted from Table 11.4 at the bottom of the page.

[7] One of the arguments revolves around the fact that the original tonal hierarchy profiles were based on data from a subset of contextual conditions in which the most stable tones occurred more often than the unstable tones. It appears that Krumhansl and Kessler selected those conditions that were most effective in inducing a sense of key, and it is not surprising that these contexts were precisely those that emphasized stable tones. Butler's question of whether the subjects' ratings reflected their sense of key or the frequency of occurrence of the stable tones seems somewhat academic.

Haskins Laboratories Status Report on Speech Research
1992, SR-109/110, 283-288

Toward an Emancipation of the "Weaker Sense"*

A review by Bruno H. Repp

Listening: An Introduction to the Perception of Auditory Events by Stephen Handel. Cambridge, MA: MIT Press, 1989.

Auditory Scene Analysis: The Perceptual Organization of Sound by Albert S. Bregman. Cambridge, MA: MIT Press, 1990.

Cognitive Foundations of Musical Pitch, by Carol L. Krumhansl. Oxford University Press, 1990.

Auditory perception has always been the stepchild of psychology. The rapid advances of computer technology have, if anything, further increased the hegemony of the visual sense: The prototypical computer combines stunning graphic capabilities with a primitive sound inventory, and it seems useless without a monitor, whereas a loudspeaker can easily be dispensed with. The optic display capabilities of computers are utilized widely in psychological research, their sound-generating capabilities only by a few specialists. Even in those branches of psychology that ostensibly deal with audible things, such as psycholinguistics and psychomusicology (not to speak of their ancient armchair relatives, linguistics and musicology), theory and experimentation are commonly based on visual representations of the objects under study. Books on auditory subjects (such as the three reviewed here) usually have plenty of figures, but no sound sheets.

It comes rather as a shock, therefore, when Stephen Handel opens his book with the confession that "In our culture, I would much prefer to be blind than to be deaf" (p. 1). With this simple but soon convincing statement, he reminds us of the unique importance of the auditory sense to our life experience: More than vision (but rather like the tactile sense, which is much more limited in range), audition keeps us "in touch" with our environment. Moreover, it is the basis of the two most important systems of human communication: speech and music. If televisionhad no sound, it would never have edged out radio as the most popular medium of news and entertainment.

Preparation of this review was supported by NIH Grant RR-05596 to Haskins Laboratories.

Psychological Science, 2, 382-386.

Much of the psychological research on hearing in recent decades has employed simple sounds and sophisticated psychophysical methods; this "psychoacoustics" continues to thrive in a somewhat segregated fashion in many laboratories and in the pages of *The Journal of the Acoustical Society of America*. Another significant area in auditory research is speech perception, which is even more of a segregated specialty, called "speech science" (or "speech technology") as soon as some application is in sight, and otherwise largely associated with work done at Haskins Laboratories and the reactions of others to it. Most of the speech perception work has been at the fringe of psychoacoustics, with speech sounds being taken apart into their smallest components until they ceased to be speech and researchers felt on familiar territory again. Much the same can be said about research on music perception, a large part of which has been concerned with pitch, duration, and loudness.

At the same time, a few small rivulets began to flow beside the mainstream of reductionistic auditory research (itself a minor tributary to the St. Lawrence of largely vision-based psychology). James Gibson, in his influential discussions of ecological principles of perception, had relatively little to say about audition (though there is a chapter in Gibson, 1966), but his emphasis on environmental objects and events, and on the perceiver's attunement to systematically structured physical media, spread seeds which germinated during the 1970s. In the 1980s, Carol Fowler emerged as a champion of an ecological perspective on speech perception (see Fowler, 1986), and James Jenkins wrote a stimulating chapter presaging a science of ecological acoustics (Jenkins,

1985). Albert Bregman's research program on auditory organization had been under way for some time and yielded a steady stream of research reports, with occasional contributions from others (e.g., Kubovy, 1981), applications to speech perception (e.g., Darwin, 1984), and a lively counterpoint of inventive experiments from Richard Warren's laboratory (e.g., Warren, 1984). Music psychology, a relatively obscure enterprise through the 1970s, suddenly gained momentum through publications such as the book edited by Deutsch (1982), Roger Shepard's (1982) influential article on pitch structures, and the extensions of his work by his former student, Carol Krumhansl.[1]

The three books reviewed here reap the harvest of these developments. One of them, Handel's *Listening*, is a broad introduction to the perception of auditory events, with special attention to speech and music. Bregman's *Auditory Scene Analysis* focuses more narrowly on the perceptual organization of simple sound patterns, but treats this topic expansively, with the author's own ideas and research at the center of attention. Krumhansl's *Cognitive Foundations of Musical Pitch* is even more specialized in that it summarizes the author's research since the late 1970s, with only brief digressions into related literature. It is also considerably more succinct than the other two tomes, and it does not share their overt ecological orientation, being squarely in the tradition of cognitive psychology. What all three books have in common is excellence.[2]

AUDITORY EVENTS

Handel's book begins with a detailed but very readable introduction to the physics of sound production, presented without mathematical formulas but with many illustrations. This is followed by a chapter on sound propagation in the environment, by two chapters dealing specifically with sound production by musical instruments and by the human vocal tract, and by a chapter summarizing acoustic (and, very briefly, perceptual) commonalities between speech and music. The remaining chapters, which constitute roughly two thirds of the book, deal with issues of perception. The first of these chapters is on auditory stream segregation and lucidly overviews a topic treated in much more detail in Bregman's book. Chapter 8, "Identification of Speakers, Instruments, and Environmental Events", is particularly valuable in pointing out the common aspects of these important activities, which have been given less research attention than they deserve. Chapter 9 deals primarily with categorical perception and context effects, with the focus on speech. Under the unfortunately misleading title, "Grammars of Music and Language", the next chapter deals almost entirely with music, particularly pitch structures, anticipating Krumhansl's more detailed treatment. (Was an earlier section on linguistic grammar deleted at the last minute?) The following chapter on rhythm is more balanced and provides a very useful discussion of music in juxtaposition with prosodic aspects of speech. The final chapter, somewhat surprisingly, is on auditory physiology, but summarizes what is known about the auditory processing of complex sounds and speech, so that it ties in well with the general thrust of the book. A brief epilogue points out two aspects that were neglected in the book: the role of the listener's expectations and knowledge, and a characterization of the experience of listening.

Handel's book contains a wealth of information, presented accurately, in simple prose with numerous instructive examples. It brings together, often for the first time, topics that have been treated in articles scattered through the research literature, and it provides a coherent perspective on them. The writing is modest, thoughtful, and balanced; there is no dogma or strident criticism, nor any oversimplification of complex issues. Handel always shows a healthy respect for the complexity of natural phenomena, and he inculcates the same attitude upon the receptive reader. As Albert Bregman says on the book jacket, "*Listening* is obviously the work of a master teacher."

AUDITORY SCENE ANALYSIS

Bregman's own book, *Auditory Scene Analysis*, is narrower in scope than Handel's but probes the topic in much greater depth. At 773 pages surely one of the heftiest monographs ever published in psychology, it rests heavily on Bregman's own research since the late 1960s and on the contributions of a few other scientists working in the same area. Its leisurely, narrative style at times gives it the quality of a historical or philosophical treatise. In a very real sense, Bregman serves as the historian of his own ideas and research. One rarely gets such an intimate view of a scientist's mind at work, nor such a comprehensive picture of personal observations, experimental explorations, and alternative interpretations. Bregman invites the reader to join him on his intellectual journey, and I, at least, found the book difficult to put down. If Handel's book shows a master teacher at work, then this is the product of a master thinker.

The term "auditory scene analysis" was coined by Bregman to refer to the process of organizing

complex auditory input into internally coherent "streams" or auditory objects. He distinguishes two classes of such processes: "primitive" and "schema-based" stream segregation. The book deals primarily with primitive processes, which do not depend on a listener's domain-specific knowledge. Bregman believes (although he acknowledges that further research is needed) that primitive scene analysis segregates auditory events before they are interpreted with reference to learned "schemas," as in listening to speech or music.

Following an introductory chapter, more than half of the book is taken up by Chapters 2 and 3, which deal with sequential (temporal) and simultaneous (spectral) integration/segregation, respectively. Chapter 2 introduces the now well-known phenomenon of auditory stream segregation and the seminal work of van Noorden (1975)—surely the most cited unpublished dissertation in the field—, and proceeds to discuss exhaustively what is known about the various factors that influence the perceptual grouping of acoustic elements. Chapter 3 discusses the factors that cause simultaneous tones to fuse into a single percept or to be perceived as separate pitches or timbres. It covers a good deal of more traditional psychoacoustics (such as pitch perception, binaural fusion, masking, etc.), but Bregman never strays very far from his own research and brings in the findings of others primarily to illuminate or supplement the story of his enterprise.

The reader who has been persistent enough to plow through these two enormous but fascinating chapters, each a small book in itself, is faced with five additional, shorter chapters. Chapter 4, "Schema-Based Integration and Segregation" is shorter because Bregman's goal is to distinguish and separate knowledge-guided processes from primitive auditory scene analysis, and to keep the focus on the latter. Perhaps the most important theoretical argument of the book is that primitive scene analysis is independent of acquired knowledge, though what has been divided by scene analysis can sometimes be recombined into a higher-level (schema-based) unit. In the popular jargon of contemporary cognitive science (which Bregman studiously avoids), primitive scene analysis is modular and noninteractive. Chapters 5 and 6 deal with auditory organization in music and speech perception, respectively. Again, these discussions focus on the role of primitive scene analysis, not on the perceptual consequences of the categories and structures specific to each system. Thus they address such basic questions as "What makes a melody hang together?" and "What makes the different voices in polyphonic music distinct?", or in speech, "Why are the sounds of speech perceived as a coherent stream?" and "How do we separate several simultaneous voices from each other"? The parallel nature of these questions in music and in speech reflects the universality of auditory scene analysis. Music- and speech-specific knowledge is considered a nuisance factor from the perspective of this book, which treats music and speech as pure sound. This may disappoint some musicologists and linguists among the readers, but Bregman should not be blamed for saying little about topics that his book is not about; rather, the rigor of his approach must be admired, for there is a continuous temptation to elevate (or, rather, reduce) knowledge-based processes to the status of auditory primitives, particularly in the case of speech.

Chapter 7 presents a relevant case study. Under the heading of "The Principle of Exclusive Allocation in Scene Analysis", Bregman discusses the phenomenon of duplex perception, an instance in which the principle is violated (i.e., the same sound appears to be heard as part of two different streams). In fact, Alvin Liberman and his collaborators at Haskins Laboratories claim that speech schemas (Bregman's term) override and even "pre-empt" auditory scene analysis (see, e.g., Liberman & Mattingly, 1989). Bregman discusses evidence to the contrary. Still, the issue remains somewhat unresolved at the end of the chapter, which is the most difficult and the least definitive in the book. The last chapter. "Summary and Conclusions: What We Do and Do Not Know about Auditory Scene Analysis" condenses the book's contents onto 65 pages, much for the benefit of readers who just want to get the gist of it. Here and throughout the volume, Bregman's honesty in acknowledging unresolved questions and missing empirical evidence is exemplary. There are many leads for future research to be done, and Bregman's accomplishment is made all the more impressive by his careful delineation of its current limits. This book will stand as an important milesteone in the history of 20th century psychology, as well as an inspiring human document.[3]

PITCH STRUCTURES

With Krumhansl's monograph we enter a different world, yet one that dovetails nicely with Bregman's and especially with Handel's introduction. Krumhansl is concerned with some of the

schema-guided processes in music perception that exceeded the scope of Bregman's book, specifically the relationships among the pitches of the Western tonal system. The monograph is a natural outgrowth of Krumhansl's exceptionally systematic and coherent research program, which is almost unique in the burgeoning field of music psychology.

Lucid and organized throughout, Krumhansl's writing lacks the old-fashioned charm of Bregman's meandering thoughts. Instead, there is a crystalline quality to her orderly designs and structural representations. Clearly, her most distinctive achievement is in the domain of sophisticated quantitative analysis. As one of Roger Shepard's most brilliant students in the 1970s, she absorbed the multidimensional techniques pioneered by her mentor and proceeded to apply them to musical problems in an imaginative and revealing way. Despite the formal complexity of these analyses, she makes the results always easy to grasp, with the help of many illustrations which are an essential part of the methodology. Having accorded Handel and Bregman master status, without intending to stereotype them in any way, I regard this book as the work of a master analyst.

Only a very brief summary of the contents can be given here. To convey the full flavor and elegance of Krumhansl's research, a much longer precis would be necessary, which will appear elsewhere (Repp, in press). Krumhansl's primary experimental technique is a probe task in which a musical context (a melodic fragment, sequence of chords, or excerpt from a composition) is followed by a probe tone or chord, whose adequacy as a continuation of the preceding context the listener is to judge. Probe elements are sampled exhaustively from a fixed set (such as the 12 tones of a scale), and a profile of average ratings across these elements is obtained. The autocorrelation matrix of this profile, which represents the similarities of the rating profiles for the same probe elements in the context of all different keys, is subjected to multidimensional scaling, which results in a spatial representation of keys similar to such maps constructed intuitively by musicologists. The ingenious part of the methodology begins when the key rating profiles established in the initial experiments are used as diagnostic tools for determining the perceived key following some arbitrary context. Thus Krumhansl devises a key-finding algorithm based on the frequency distribution of the most recent notes and their correlations with all possible key profiles; by presenting probe tones after some musical excerpt, she determines which (if any) key is dominant at that point by finding the prototypical key profile that most closely resembles the obtained rating profile; and in the most advanced application of these procedures, she traces the listener's changing sense of key through a modulating sequence of chords by obtaining a probe tone profile after each chord, correlating each of these profiles with the standard key profiles, and finally mapping these relationships into a multidimensional space, where they trace a modulatory path among the various keys (represented as points in the space). High-wire acts such as these are complemented by results from simpler memory tasks and other studies in the literature.

The central chapter of the book is Chapter 6, which defines three principles of tonal stability, and their effects on the perceived relations between tones. Tonal stability is the central concept of the book, and indeed of traditional music theory; it refers to the fact that, in tonal music, there is a hierarchy of pitches, such that one pitch (the key-defining pitch or tonic) is most preferred or most important or most representative—in other words, most stable—at any given time, a second pitch (the dominant) is preferred next, and so on. Krumhansl's three principles, then, are (in simplified language): A stable tone seems more similar to itself (e.g., is remembered better) than an unstable tone; two tones seem more similar to each other if either of them is stable; and two tones of unequal stability seem more similar when the unstable tone precedes the stable one. A variety of evidence supports these statements, which parallel predictions made with regard to prototypicality in many other areas of cognitive psychology. Krumhansl's work indeed falls squarely into mainstream cognitive psychology and should contribute significantly to making music psychology seem part of this larger enterprise.

THREE FIELDS ON THE MOVE

It is perhaps appropriate to conclude this review with some musings on the current state of three fields of research that are addressed by the books reviewed (ecological acoustics, speech perception, and music perception), and on the influence the books might have on research in the 1990s. As it happens, the three fields named are at rather different stages of development: one nascent, one burgeoning, and one temporarily stagnant. These impressions are subjective, of course, and depend in large measure on where I draw the boundaries of these domains of inquiry.

Ecological acoustics. Under this rubric I would consider studies that deal with the analysis and perception of information in complex sounds other than the message elements of speech or music—information that helps us identify individuals, objects, and events in our environment. (Ultimately, of course, the "ecologically valid" study of speech and music as gestural events must be included, too.) Under this definition, Bregman's work is merely a prolegomenon to an ecological acoustics, though an essential one. Handel's Chapter 8 ("Identification of Speakers, Instruments, and Environmental Events") is most pertinent, inasmuch as speaker and instrument identification are not really linguistic or musical activities. The literature on human speaker identification is relatively small (much smaller than that on automatic speaker recognition), and most of it originates in Europe. Research on instrument identification is almost nonexistent. The related topic of the acoustic expression and perception of emotion in speech and music is likewise under-researched, with Klaus Scherer's work on speech standing as a single beacon in the desert (see Scherer, 1986). Handel discusses Warren and Verbrugge's (1984) study of breaking and bouncing events, which is a prototype for ecological acoustics research built on Gibsonian premises, but little has happened since except for a few isolated studies on seemingly exotic topics including "chilling" sounds (Halpern, Blake, & Hillenbrand, 1986), hand clapping (Repp, 1987), and the sounds of kitchen pans struck with mallets (Freed, 1990). However, those who doubt the potential significance of studies in ecological acoustics may be converted by reading Tom Johnson's (1984) still unpublished dissertation on doctors' perception of human heart beats, a brilliant foray into real-world relevance. The message of all these studies is that we hear not just sounds but, through their structure, environmental happenings and organisms in action. Hopefully, Handel's book will stimulate more research on how we use our ears to perceive actions and events—how we hear the world.

SPEECH PERCEPTION

Research on speech perception began at Haskins Laboratories in the early 1950s and largely remained dependent on the technology available there for the next two decades. Then, with computers getting smaller and cheaper, and with software replacing hardware synthesizers, other laboratories got into the business. Much of that research, however, remained methodologically and conceptually dependent on the Haskins research: Nearly everyone tried to support, refute, or extend the claims of the Haskins researchers. (Among the few significant exceptions, Richard Warren's consistently original—though psychoacoustically tinged—contributions are especially noteworthy; see, e.g., Warren, 1982, 1984.) The 1970s and early 1980s were fertile years for speech perception research, with several popular paradigms being milked dry and lively arguments going back and forth. Now these activities seem to have slowed down, very much in proportion to the decline of speech perception research at Haskins Laboratories, where most of the effort is nowadays directed to speech production. The older generation of speech perception researchers has reached retirement age, and many of the younger (now middle generation) protagonists of the 1970s have turned to different topics or tend to publish less, with only a few die-hards continuing to suck on the dry teats of their superannuated paradigms. There seems to be a general lack of intellectual ferment in the field.

Was speech perception research (as defined rather narrowly here) just a historic episode? Did Bregman in his chapter on duplex perception capture the last, already somewhat peripheral controversy? Perhaps, as far as the dominant and unifying (or, rather, constructively divisive) role of Haskins Laboratories is concerned. It may take a while before new ideas develop and strong voices emerge to put them forward. Handel's book will make only a minor contribution here; what is needed is a coherent body of research that makes a point, comparable to what Krumhansl and Bregman have to offer. The ideas of the most innovative theorist in recent years, Carol Fowler, hold much promise but have not yet resulted in a critical mass of empirical findings. Meanwhile, of course, there is a rich matrix of related research activities in experimental phonetics, speech science, speech technology, and psycholinguistics, which are not experiencing a similar recession and may provide hotbeds for new directions in speech perception research.

MUSIC PERCEPTION

Music psychology, and particularly research on music perception, is on the rise. One of the prime movers is Diana Deutsch who has earned the socio-scientific triple crown by editing the first modern collection of articles on the subject (Deutsch, 1982), founding the journal *Music Perception* in 1983, and by recently establishing the Society for Music Perception and Cognition, in addition to be-

ing a fertile and original researcher at the psychoacoustic end of the music spectrum. A number of other significant books have appeared in recent years, of which Sloboda's (1985) is the most original, and music-related conferences abound. There is a rapidly increasing pool of talented young investigators, each of whom quickly seems to find a niche in the vast territory offered by musical questions and phenomena. Interdisciplinary conferences bring psychologists together with music technologists, musicologists, composers, and performers—some sceptical, to be sure, but all eager to exchange ideas and explore new avenues. Electronic instruments, sophisticated software, and MIDI systems offer new and exciting possibilities for research and practice. As an added special touch, a shared love for music unites scientists of very different theoretical persuasions: The fact that music gives aesthetic pleasure and spiritual sustenance is never far from their minds and frequently invades their discussions, whereas speech researchers, for example, rarely think of drama or poetry in connection with their work.

Krumhansl's book rides the crest of a wave to which she herself contributed significantly. The book serves mainly to bring her work to the attention of those who have not been following her progress in the journals, and it is admirably suited for that purpose. The research itself, of course, has led and influenced the field for some time, and also has aroused some controversy (Butler, 1989; Krumhansl, 1990; Butler, 1990)—a healthy sign of a science's vitality (cf. Hull, 1988). Unlike Bregman's life work, which has the quality of a fortress under construction, with open doors but numerous escape routes, all thoroughly explored in advance of any possible attack, Krumhansl's work, with its carefully planned design, its built-in dependencies among experiments, and its strong reliance on one particular methodology, appears much more vulnerable and transparent, more like a contemporary office building in a historic neighborhood. It remains to be seen whether her constructs and methods can withstand critical onslaught. Meanwhile, her book is required reading for anyone interested in the contemporary psychology of music, as indeed are the other two volumes reviewed here. In concert, this admirable trio should convince anyone that auditory perception is worthy of much more attention by psychologists than it has received in the past.

REFERENCES

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, 6, 219-242.

Butler, D. (1990). Response to Carol Krumhansl. *Music Perception* 7, 325-338.

Darwin, C. J. (1984). Auditory processing and speech perception. In H. Bouma & D. G. Bouwhuis (eds.), *Attention and performance X. Control of language processes* (pp. 197-209).

Deutsch, D. (Ed.) (1982). *The psychology of music*. New York: Academic.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*, 87, 311-322.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Halpern, D. L., Blake, R., & Hillenbrand, J. (1986). Psychoacoustics of a chilling sound. *Perception & Psychophysics*, 39, 77-80.

Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.

Jenkins, J. J. (1985). Acoustic information for objects, places, and events. In W. H. Warren & R. E. Shaw (eds.), *Persistence and change. Proceedings of the First International Conference on Event Perception* (pp. 115-138). Hillsdale, NJ: Erlbaum.

Johnson, T. L. (1984). *Expertise in cardiac auscultation: Perception of relative intensities and timing of second heart sound components*. Unpublished dissertation, University of Minnesota.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, 7, 309-324.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55-98). Hillsdale, NJ: Erlbaum.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243, 489-494.

Noorden, L. P. A. S. van (1975). *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Eindhoven University of Technology.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America*, 81, 1100-1109.

Repp, B. H. (in press). Review of *Cognitive foundations of musical pitch* by Carol L. Krumhansl. *American Journal of Psychology*.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305-333.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Clarendon Press.

Warren, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon Press.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, 96, 371-383.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 704-712.

FOOTNOTES

[1] Naturally, these are just selected highlights which happened to leave a strong impression on me.

[2] What is missing from my shelf is a research monograph on speech perception from a cognitive or ecological perspective.

[3] A useful supplement to the book would have been a soundsheet or CD illustrating the auditory phenomena discussed in the book. Some years ago, Bregman produced a cassette with such demonstrations and distributed copies to colleagues; I understand that copies can still be obtained by sending $5.00 to him at McGill University.

Appendix

SR #	Report Date	NTIS #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329	**
SR-97/98	January-July 1989	PB 90-121161	ED32-1317
SR-99/100	July-December 1989	PB 90-226143	ED32-1318
SR-101/102	January-June 1990	PB 91-138479	ED325-897
SR-103/104	July-December 1990	PB 91-172924	ED331-100
SR-105/106	January-June 1991	PB92-105204	ED340-053
SR-107/108	July-December 1991	PB92-160522	ED344-259
SR-109/110	January-June 1992		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned